

TAGUNG

Hochschulranking
Aussagefähigkeit, Methoden, Probleme

DOKUMENTATION

Herausgeber

Detlef Müller-Böling

Stefan Hornbostel

Sonja Berghoff



Centrum für
Hochschulentwicklung



Verlag Bertelsmann Stiftung

Detlef Müller-Böling, Stefan Hornbostel, Sonja Berghoff

Hochschulranking

Aussagefähigkeit

Methoden

Probleme

Wissenschaftliche Tagung

Berlin 25./26. September 2000

Veranstalter: CHE Centrum für Hochschulentwicklung

Verlag Bertelsmann Stiftung

Gütersloh 2001

Die Deutsche Bibliothek – CIP –Einheitsaufnahme

Ein Titeldatensatz für diese Publikation ist bei der Deutschen Bibliothek erhältlich.

©2001 Bertelsmann Stiftung, Gütersloh

Verantwortlich: Detlef Müller-Böling

Umschlaggestaltung: Frank Wellenbrink, P&P

Druck und Bindung: Setter Druck und Beschriftung, Rietberg

ISBN 3-89206-575-5

www.bertelsmann-stiftung.de/verlag

Inhalt

Detlef Müller-Böling	
Einführung	5
Stefan Hornbostel	
Hochschulranking: Beliebigkeit oder konsistente Beurteilungen? Rankings,Expertengruppen und Indikatoren im Vergleich	7
Helmut Kromrey	
Informationsgehalt und Validität von Studierendenbefragungen als Evaluation	43
Uwe Jensen	
Bootstrap-Rangkonfidenzintervalle für Hochschulranglisten.....	65
Christian R. Bayer	
Vergleich zwischen deutschen und amerikanischen Ranking-Studien.....	91
Hans-Dieter Daniel	
Was bewirken Hochschulrankings?	121
Jürgen Güdler	
Drittmittel als Indikator für Forschungsleistung. Analysen zum DFG- Bewilligungsaufkommen	125
Ulrich Schmoch	
Aussagefähigkeit von Patentanalysen zur Bewertung von Hochschulen	143
Michael Weegen	
Informationssystem Studienwahl und Arbeitsmarkt – ISA	157
Autoren	167
Das CHE Hochschulranking - Untersuchungen und Bibliographie	169

Hochschulranking - Aussagefähigkeit, Methoden, Probleme

Eine Einführung

Detlef Müller-Böling

Seit Ende der achtziger Jahre die ersten Hochschulrankings in Deutschland erschienen, sind sie umstritten. Nach den ersten Anfängen haben sie sich jedoch methodisch erheblich weiterentwickelt. Befragungsergebnisse von Studierenden wie Professoren fließen heute ebenso ein wie Fakten, basierend auf statistischen Angaben, Patentanalysen und bibliometrischen Messungen. Doch wie ist es bestellt um ihre Aussagekraft? Ist ihre Wirkung bei der Studienplatzwahl nachweisbar? Welche methodischen Fallstricke birgt der Vergleich hochschulischer Leistungen in Forschung und Lehre? Von der Analyse von Patentaufkommen und Drittmitteln zur Beurteilung von Forschungsleistungen über statistisch-methodische Neuerungen, Evaluation mittels Befragungen bis zur Information über den Arbeitsmarkt: Die in diesem Tagungsband zusammengefassten Beiträge spiegeln die breite Palette methodischer Herausforderungen bei der Erstellung fundierter Hochschulrankings. Beispielhaft steht in den meisten Beiträgen das umfassendste deutsche Ranking, das jährlich aktualisiert vom CHE (in Zusammenarbeit mit dem stern) veröffentlicht wird (www.dashochschulranking.de).

Die öffentliche Debatte um Rankings wird weitgehend empiriefrei geführt. Das CHE hat kritischen Fachexperten Datenmaterial aus dem Hochschulranking zur Verfügung gestellt, um den Einwänden nachzugehen. Kritik ist Ausdruck wissenschaftlicher Auseinandersetzung. Das CHE-Hochschulranking will ein wissenschaftsfundiertes Ranking sein. Von daher ist Kritik explizit erwünscht und in diesem Band dokumentiert. Dies ist keine Festschrift für das CHE-Hochschulranking, sondern eine kritisch-konstruktive Auseinandersetzung.

Die Beiträge zeigen, dass es im Detail noch vieles zu feilen gibt. Im großen und ganzen zeigt sich aber, dass das Hochschulranking auch von seiner methodischen Anlage auf dem richtigen Weg ist. So geht es in den Beiträgen letztlich nicht um eine Fundamentalkritik, sondern um eher um die Frage im Detail, wie sich der Informationsgehalt noch weiter steigern lässt.

Die Beiträge basieren auf einer Tagung. Aus den Vorträgen und den Diskussionen haben wir konkret folgende Konsequenzen gezogen:

1. Ausweitung der betrachteten Fächer bzw. Studienbereiche

Das Informationsbedarf unter potentiellen Studienanfängern ist enorm, die Versorgung mit Informationen nur spärlich. Ab 2001 werden mit dem CHE Hochschulranking 25 Fächer abgedeckt sein. Diese Informationen werden ab 2002 in dreijährigem Turnus wiederaufgefrischt und mit Aussagen zur Entwicklung der Fachbereiche angereichert.

2. Zugang zum Datenmaterial für Zielgruppe

Das im Internet online und kostenlos vorhandene Datenmaterial ist zu komplex, als dass ein einfacher Zugang möglich wäre. Dadurch werden viele wichtige Informationen von Informationssuchenden gar nicht erst entdeckt. Durch eine Überarbeitung der Internet-Darstellung, die insbesondere eine stärkere Führung enthalten wird, erhoffen wir uns höhere Zugriffszahlen und ein besseres Ausschöpfen der reichlich vorhandenen Informationen.

3. Identifikation verschiedener Studierendentypen

Offensichtlich sind Studienanfänger keine homogene Masse sondern legen verschiedene Kriterien zur Auswahl des Studienortes an. Um das Matching zwischen gewählter Hochschule und Studienanfänger zu verbessern, haben wir Studierendentypen charakterisiert, für die wir die ‚besten‘ Hochschulen zusammenstellen konnten.

4. Arbeitsmarkt

Sicherlich ist der Übergang in den Arbeitsmarkt ein wesentlicher Aspekt, der unserem Studienführer noch fehlt. Neben eigenen Aktivitäten z. B. der Aufnahme von Fragen zu diesem Bereich in unsere aktuellen Erhebungen oder Alumni-Befragungen, planen wir hier eine Zusammenarbeit mit ISA.

5. Methodisches

Methodischen Neuerungen werden wir uns nicht verschließen, auch neue Verfahren wie das Bootstrapping werden sicher eine Chance erhalten. Besonders wichtig sind Methoden zur sinnvollen Verdichtung der schier Masse von Daten, ohne dabei wesentliche Informationen zu verlieren.

6. Wirkungsanalysen

Welche Wirkung hat unser Ranking? Im Hochschulranking 2002 werden erstmals Fächer in einer zweiten Runde untersucht. Dann wird es Aussagen geben zu Veränderungen - von denen einige sicherlich auch auf das CHE-Ranking zurückzuführen sein werden.

Hochschulranking: Beliebigkeit oder konsistente Beurteilungen?

Rankings, Expertengruppen und Indikatoren im Vergleich

Stefan Hornbostel

Vor mehr als zehn Jahren wurde in Deutschland das erste Hochschulranking veröffentlicht. Damals eher belächelt oder beschimpft, wird es heute, sehr viel ernster genommen (vgl. Hornbostel 1999a). Eine wachsende Zahl von Studienanfängern zieht inzwischen Rankingergebnisse für ihre Studienentscheidung zu Rate, wenngleich die wichtigste Determinante für die Standortwahl immer noch die Nähe zum heimatlichen Herd ist (vgl. Daniel in diesem Band, Lewin, K. u.a. 1997). Die Rankings selbst haben sich in den vergangenen 10 Jahren erheblich verändert. Dazu muss man nur einmal die jüngste CD des CHE Studienführers mit dem ersten SPIEGEL Ranking vergleichen. Der kritischen Auseinandersetzung mit Rankings kann man eine solche Entwicklung allerdings nicht attestieren. Ein Großteil der Kritik wiederholt stereotyp die Argumente und Verdachtsmoment, die schon vor zehn Jahren vorgebracht wurden.

Eine typische Zusammenfassung der Debatte um Hochschulrankings liest sich folgendermaßen: „Alle sind methodisch bedenklich oder sogar falsch. Die eigentlich propagierten Ziele werden durch die Fragestellungen nur teilweise erreicht. Die Rankingmodelle in Deutschland sind noch immer auf subjektiven Reputationsvergleichen aufgebaut. Im Gegensatz dazu werden die Ranglisten in den USA, die weitaus mehr Tradition als in Deutschland haben, auf Basis von objektiven Kennzahlen erstellt. Die Datenerhebungen selbst beinhalten viele Fehlerquellen wie zu wenig Befragte, falsche Auswahl der Befragten oder fehlerhafte Antworten. Wegen der Unterschiedlichkeit der verwendeten Quellen / Befragten, Fragestellungen, Evaluationskriterien und Bewertungsmaßstäbe sind die Rankings eigentlich nicht miteinander vergleichbar“ (Ott 1999: 320). Diese Liste lässt sich fortsetzen: Gemessen werde mit Hochschulrankings „well being“ sagen die einen (Scheuch 1990), Studierende in den Metropolen seien kritischer als in kleineren Orten, sagen die anderen (Meinefeld 2000) und weisen darauf hin, dass es für Bewertung keine allgemeinverbindlichen Maßstäbe gäbe.

Es herrscht heute zwar weitgehend Einigkeit darüber, dass Evaluation und Ranking notwendig, zumindest aber unumgänglich sind, sobald es aber um tatsächlich durchgeführte Rankingstudien geht, fallen die Einschätzungen über den Ertrag und die Folgen weit auseinander. Während die einen loben, dass Stärken und Schwächen sichtbar, Orientierungsgrundlagen für die Nachfrager hochschulischer Leistungen geschaffen würden, Blockaden und Verkrustungen im Selbststeuerungssystem aufgebrochen oder gar effizienz- und leistungssteigernde Allokationsmechanismen möglich seien, wird von den anderen vorgetragen, dass günstigstenfalls eine Art Infotainment mit einer Fülle widersprüchlicher Informationen entstanden sei oder eine Vergeudung von Zeit und Ressourcen, schlimmstenfalls jedoch ein methodisch fragwürdiges Instrumentarium, das möglicherweise sogar Fehlinformationen und Irreleitungen begünstige oder ein Feigenblatt für weitere Etatkürzungen liefere. Kurz: „Ranking ist oft ein Synonym für Unsinn“ (Süllwoll 1997). Es ist bedauerlich, dass die Auseinandersetzung mit Rankingstudien nach wie vor überwiegend aus Spekulationen und Vermutungen gespeist wird, denn inzwischen liegt eine Fülle von Daten vor, die eigentlich dazu einlädt, sich einmal in konstruktiv-kritischer Weise mit Hochschulrankings auseinander zusetzen.

Der folgende Beitrag versucht, die wichtigsten Einwände gegen Rankingstudien aufzunehmen und mit empirischen Befunden vor allen Dingen aus den CHE Studienführern zu konfrontieren.

1. Einwand: Studierendurteile sind eine unzuverlässige Informationsquelle

Dieses Argument taucht seit Anbeginn der Rankingdebatte in den verschiedensten Versionen auf: es reicht von der Behauptung, dass Studierendurteile allenfalls ein Wohlfühlindikator seien, über die Annahme, Studierendurteile seien vom Aspirationsniveau oder der Eingangsqualifikation der Studierenden abhängig, von Alter, Geschlecht oder der Studienfinanzierung, von den Vergleichsmöglichkeiten mit anderen Hochschulen usw. Im Kern handelt es sich um zwei unterschiedliche Argumente: Einerseits um die Vermutung, Studierende seien zu sachadäquaten Urteilen nicht in der Lage, dabei wird gern darauf verwiesen, dass Professoren häufig gerade jene Hochschulen empfehlen, die von den Studierenden negativ beurteilt werden. Andererseits um ein zweites – damit verwandtes – Argument, das behauptet, dass die *studentischen* Urteile zwar durchaus kompetent seien, aber das Urteil selbst je nach Erwartungen und Voraussetzungen unterschiedlich ausfalle, so dass am Ende nicht die Qualität des Lehrangebotes, sondern Charakteristika der Urteilenden gemessen werden.

Beide Einwände sind empirischer Prüfung zugänglich. Um zu prüfen, ob Studierende überhaupt sachliche Urteile abgeben können, benötigt man eine andere, davon unabhängige Expertise, zum Vergleich. Es liegt nahe, hier die Einschätzungen der Professoren zu Rate zu ziehen, auch wenn beide Gruppen wahrscheinlich eine etwas unterschiedliche Perspektive auf die Lehrsituation haben, schließlich sind die einen Anbieter von Lehre, die anderen Konsumenten. Wären die Kritiken richtig, dass Studierende gerade die hervorragende akademische Ausbildung negativ bewerten und stattdessen nur Nestwärme honorieren, dann müssten die Ranglisten auf den Kopf gestellt werden, wie Scheuch (1990) das einmal formulierte. Da die Professoren sicherlich nicht unter den Verdacht fallen, Nestwärme zu honorieren, müssten die Urteile beider Gruppen negativ korreliert sein.

Wie die Tabelle 1 zeigt, existieren negative Korrelationen zwischen Studierenden- und Professorenurteilen überhaupt nicht, stattdessen zeigt sich eine nach Beurteilungsgegenstand und Fach unterschiedlich hohe Übereinstimmung zwischen beiden Urteilen. Einig sind sich beide Gruppen weitgehend in der Beurteilung relativ einfacher Sachverhalte (wie etwa der Raumsituation). Auf einzelnen Beurteilungsdimensionen und auch bei der Gesamteinschätzung der Studiensituation fallen die Übereinstimmungen fachspezifisch aus. Dabei ist allerdings zu berücksichtigen, dass beiden Gruppen für das Gesamturteil eine unterschiedliche Frage vorlag. Auch die Art der Übereinstimmung lässt fachspezifische Differenzen erkennen: In der Rechtswissenschaft fallen die Urteile in der gleichen Richtung aus, aber auf unterschiedlichen Niveaus (die Professoren vergeben bessere Noten als die Studierenden), in der Physik hingegen stimmen die vergebenen Noten in hohem Maße überein und bei den Architekten herrscht vergleichsweise geringe Übereinstimmung zwischen Professoren und Studierenden. Insgesamt lässt sich jedoch festhalten, dass größerer Dissens zwischen beiden Gruppen nicht die Regel ist, sondern nur an wenigen Hochschulen auftritt. Trotz der unterschiedlichen Perspektiven von Produzenten und Konsumenten kommen beide Gruppen in der Tendenz zu ähnlichen Beurteilungen.

	Rechts wissen.	Physik	Infor- matik	Bauing.	Masch. bau	Mathe- matik	Elek- tro- tech.	Archi- tektur
Index Räume	.80**	.77**	.75**	.70**	.77**	.66**	.55**	.55*
Index Bibliothek	.58**	.46**	.47**	.63**	.44*	.20	.43*	.45
Kontakt Stud.-Prof.	.55**	.60**	.28	.57*	.48*	.36*	.66**	.17
Index PC	.46**	.44**	.52**	.11	.39*	.55**	.50*	.46
Gesamturteil Lehre ¹	.51**	.42**	.40**	.52*	.35	.36*	.27	.24

Tab. 1: Korrelationen (Pearson) zwischen den Fachbereichsmittelwerten der Beurteilungen von Professoren und Studierende. Quelle: CHE Studienführer 1999 und 2000.

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Fazit: Berücksichtigt man, dass auch ausgewiesene Experten (etwa Gutachter bei der Beurteilung von Manuskripten oder Drittmittelanträgen) erheblichen Urteilsdissens aufweisen, dann deutet das Maß an Übereinstimmung zwischen Professoren- und Studierendenurteilen auf hohe Urteilskompetenz bei den Studierenden hin.

2. Einwand: Professoren empfehlen gerade die Hochschulen, die von Studierenden schlecht beurteilt wurden

In der Berichterstattung über Rankingstudien wird immer wieder darauf hingewiesen, dass Professoren andere Hochschulen empfehlen als die von den Studierenden positiv beurteilten. Wie ist eine solche Diskrepanz zu erklären?

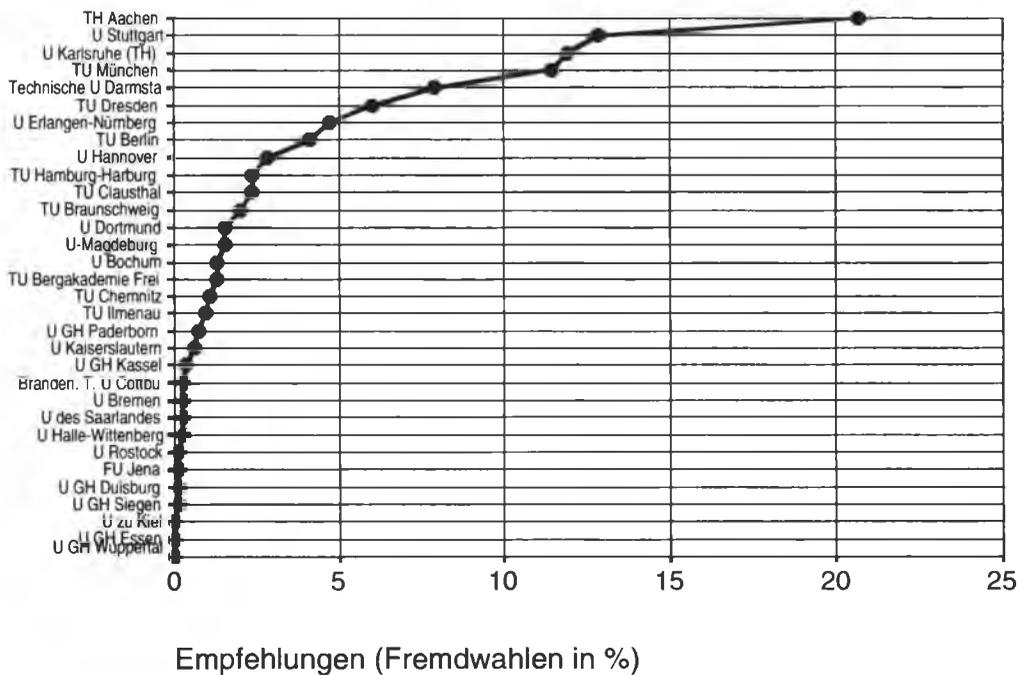
Für den obigen Vergleich wurden die Professoren gebeten, die Lehrsituation vor Ort zu beurteilen. Daneben wurden in der CHE Befragung – wie auch in anderen Rankingstudien – die Professoren um eine Empfehlung, einen „Studientipp“ gebeten (also um eine Fremdeinschätzung; vgl. dazu den Beitrag von Daniel in diesem Band). Während die ersteren Professorenauskünfte sich auf die eigenen Erfahrungen an der derzeitigen Hochschule stützen, basieren die Empfehlungen auf einer Beobachtung der Kommunikation über universitäre Leistungen. Kommuniziert werden im Wissenschaftssystem aber nicht Lehr-, sondern Forschungsleistungen (vgl. ausführlich Hornbostel 2001a).

Die Studienortempfehlungen der Professoren haben in allen Fächern die Form einer sehr schiefen und steilen Verteilung. Renommee und Bekanntheit konzentrieren sich meist auf wenige Hochschulen. Diese Verteilung ist unabhängig davon, ob „Selbstempfehlungen“ der eigenen Universität berücksichtigt oder aus der Wertung ausgeschlossen werden. Abbildung 1 zeigt einen solchen typischen Verlauf des „Professorentipps“.

¹ Die Frage für die Studierenden lautete: „Wenn Sie nun einmal alles zusammen betrachten: Wie beurteilen Sie insgesamt die Studiensituation in Ihrem Fach an Ihrer derzeitigen Hochschule?“

Die Frage für die Professoren lautete: „Wie beurteilen Sie insgesamt die Lehrsituation in Ihrem Fach an Ihrer derzeitigen Hochschule?“

Abb. 1: Hochschultipp der Professoren
Maschinenbau / Verfahrenstechnik (1999)



Für die Fremdwahrnehmung einer Hochschule bzw. eines Fachbereiches ist die Quantität und die Qualität des Forschungsoutputs wesentlich. Tabelle 2 zeigt diesen Sachverhalt recht deutlich. Sowohl die verschiedenen Größenangaben als auch die diversen Forschungsindikatoren korrelieren positiv mit der Zahl der abgegebenen Empfehlungen für eine Hochschule. Dass dabei in der Rechtswissenschaft die Drittmittelwerbungen keine signifikanten Korrelationen aufweisen, hat mit der extrem niedrigen Drittmittelaktivität in diesem Fach zu tun und mit dem Umstand, dass die Bewertungskriterien für Forschung unschärfer sind und die adressierbaren Auditorien nicht nur wissenschaftsinterner Art sind, sondern auch in die Rechtspraxis ausgreifen (vgl. Hornbostel 1999).

Betrachtet man nun die subjektiven Urteile zunächst der Professoren, zeigt sich, dass an den besonders stark empfohlenen Hochschulen die Forschungssituation vor Ort tendenziell besser bewertet wird, die Beurteilung der Lehrsituation hingegen korreliert weitaus schwächer und nicht signifikant mit den Empfehlungen der Professoren im Fach². Vergleicht man diese Beurteilung der Professoren mit den studentischen Bewertungen, lassen sich keine gravierenden Unterschiede ausmachen. Ebenso wie die Studierenden bewerten die Professoren vor Ort die Lehrsituation weitaus kritischer, als die Kollegen anderer Universitäten mit ihren Studienortempfehlungen zum Ausdruck bringen.

² Die negative Korrelation entsteht, weil hier gute Benotungen (niedrige Zahlenwerte) mit einem hohen Prozentsatz der abgegebenen Empfehlungen (hohe Zahlenwerte) korrespondieren.

	Informatik	Mathematik	Physik	Jura
Struktur				
Studierende am Fachbereich	.43 **	.43 **	.35 **	.41 **
Studierende an der Universität gesamt	.02	.30 *	.35 **	.37 *
Prüfungen je Prof. im Fach (1997)	.52 **	.23	.34 *	.47 **
Forschung				
Promotionen je Prof. im Fach (1997)	.46 **	.29 *	.36 **	.64 **
Drittmittel (1996-98) absolut	.82 **	.26 *	.58 **	.21
Drittmittel (1996-98) je Wissenschaftler	.59 **	.36 **	.25	-.04
Artikel 1996-98 im SCI (absolut)		.49 **	.73 **	
Artikel 1996-98 im SCI je Wissenschaftler		.41 **	.47 **	
Zitate je Publikation			.36 **	
Urteile				
Gesamtbeurteilung Forschungssituation (Prof.)	-.41 *	-.36 **	-.32 *	-.06
Gesamtbeurteilung Lehrsituation (Profess.)	-.22	-.20	-.08	.08
Gesamtbeurteilung Studiensituation (Studierend.)	-.01	.01	.10	.11

Tab. 2: Korrelationen (Pearson) zwischen Hochschulempfehlungen der Professoren (Prozent der Nennungen ohne Selbstwahlen) und Strukturindikatoren sowie Beurteilungen von Professoren und Studierenden

Fazit: Die Differenz zwischen Studienortempfehlung der Professoren und den Bewertungen der Lehr- und Studiensituation vor Ort entspricht einer Differenz von Fremd- und Selbstbild, wobei das Fremdbild wesentlich über die Forschungsleistungen geprägt wird. Studierende urteilen dabei nicht anders als Professoren auch. Die Studienempfehlungen der Professoren indizieren Forschungsperformanz.

Auch die Frage, ob das studentische Urteil durch Merkmale der Studierenden verzerrt wird, ist immer wieder empirisch überprüft worden (z. B. Daniel / Hornbostel 1994 und 1996). Die Daten des CHE Studienführers bestätigen zunächst einmal die bisherigen Ergebnisse. Wie Tabelle 3 zeigt, geht der ganz überwiegende Teil der aufgeklärten Varianz im studentischen Gesamturteil auf die Zugehörigkeit zu einem Fachbereich zurück und nicht auf individuelle Merkmale der Studierenden. Geschlecht und Studiendauer spielen überhaupt keine Rolle für die Beurteilung der Studiensituation. Das Lebensalter der Studierenden wirkt sich nur geringfügig aus. Das Eingangsniveau der Studierenden (hier über die Abiturnote erfasst) beeinflusst nur in der Rechtswissenschaft und der Elektrotechnik geringfügig das Urteil. Studierende mit schlechten Schulabgangsnoten beurteilen ihre Studiensituation etwas schlechter (dichotomisiert man die

Population in Schulnote ≤ 2 und > 2 , dann beträgt die mittlere Differenz in der Gesamtbeurteilung der Studiensituation in der Rechtswissenschaft allerdings nur 0,26 Skaleneinheiten).

	Maschbau	Elektrotechnik	Bauing	Architektur	Informatik	Mathematik	Physik	Rechtswiss.
Fachbereich	.72	.53	.64	.80	.72	.55	.69	.66
Forschungs-/ Praxisorientiert ³	.07	.11	.14	.02	.13	.21	.07	.07
Vollzeit- / Teilzeitstudierend ⁴	.13	.11	.11	.03	.11	.07	.09	.07
Alter	.00	.14	.01	.04	.00	.14	.08	.02
Geschlecht	.01	.00	.00	.05	.00	.00	.00	.03
Abiturnote	.04	.10	.08	.00	.03	.03	.04	.11
Job	.01	.00	.00	.00	.01	.00	.01	.00
Fachsemester	.03	.02	.03	.06	.00	.01	.02	.04
korrigiertes R ²	.19	.24	.23	.27	.21	.21	.22	.29
Valide Fälle	1931	1270	1270	957	1037	1293	1211	1307

Tab. 3: Regression (CATREG) auf das studentische Gesamturteil (abhängige Variable) - relative Wichtigkeit nach Pratt und korrigiertes R². Quelle: CHE Studienführer 1999 und 2000.

In den Ingenieurwissenschaften, der Informatik und der Physik wird, wenn man vom Einfluss des Hochschulstandorts absieht, das Urteil geringfügig vom zeitlichen Engagement der Studierenden beeinflusst. Die ca. 15% Physikstudierenden, die sich selbst als „Teilzeitstudierende“ einstufen, obwohl dieser Status de jure nicht existiert, beurteilen ihre Studiensituation insgesamt auf der Schulnotenskala von 1 bis 6 fast einen halben Punkt schlechter als ihre Kommilitonen. Auch die etwa 27% Informatikstudierenden, die sich als Teilzeitstudierende definieren, urteilen in dieser Größenordnung schlechter als ihre Kommilitonen. In diesen Fächern bieten sich für Studierende bereits im Studium attraktive Möglichkeiten auf dem Arbeitsmarkt. Möglicherweise verschiebt ein quasi-berufliches Engagement die Erwartung an Inhalte und Effizienz des Studiums. Allerdings in einer Größenordnung, die – würde man diesen Effekt herauspartialisieren – die Positionierung der Fachbereiche fast unverändert lässt.

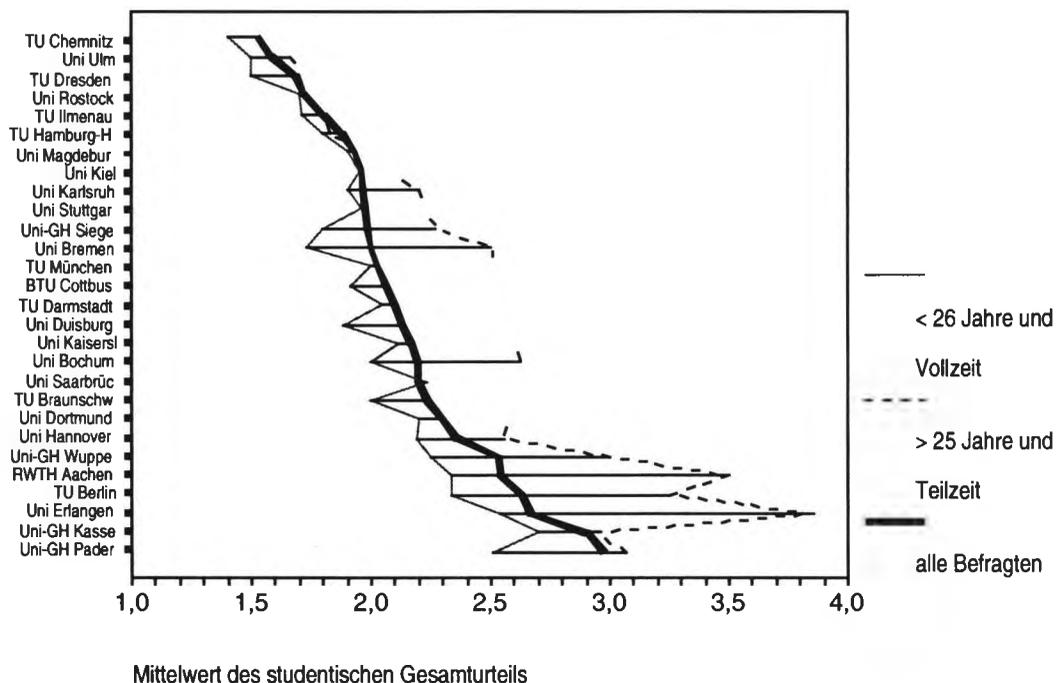
3 Aus den Fragen nach der Wichtigkeit des Forschungsbezuges bzw. des Praxisbezuges des Studiums (6 stufige Skala) wurden drei Typen gebildet: 1. Studierende, für die der Forschungsbezug wichtig ist (Skaleneinheiten 1 und 2) und der Praxisbezug gleich oder weniger wichtig ist; 2. Studierende, für die der Praxisbezug wichtig ist (Skaleneinheiten 1 und 2) und der Forschungsbezug weniger wichtig ist; 3. Studierende, für die sowohl der Forschungsbezug als auch der Praxisbezug weniger wichtig ist (Skaleneinheiten > 2).

4 Die Angaben beruhen auf einer Selbsteinstufung als Vollzeit oder Teilzeitstudierend, unabhängig vom Vorhandensein einer entsprechenden rechtlichen Regelung.

Von den 1.565 befragten Studierenden der Elektrotechnik gehören ca. 66% zum „Normaltypus“ (jünger als 27 Jahre und Vollzeitstudierend) und ca. 11% zum in der Regel besonders kritisch urteilendem Typus, der älter als 26 Jahre ist und sich als teilzeitstudierend einstuft. Letztere Studierenden urteilen im Durchschnitt ca. 0,7 Skalenpunkte schlechter (Gesamturteil) als die Studierenden aus der ersten Gruppe. Der Anteil der älteren Teilzeitstudierenden an allen Studierenden der einzelnen Fachbereiche schwankt zwischen ca. 2% und ca. 32% . Die Existenz solcher Subgruppen wird häufig als Argument gegen die Aussagekraft von Mittelwerten (vgl. Süllwold 1997) verwandt, da sich im Extremfall ähnlich große Gruppen mit diametral entgegengesetzten Urteilen an einem Fachbereich befinden. Praktisch sieht es allerdings meist anders aus, wie die folgende Abbildung 2 zeigt. In der Graphik werden die Urteilsmittelwerte der jungen Vollzeitstudierenden, der älteren Teilzeitstudierenden und aller Studierenden in der Elektrotechnik verglichen (berücksichtigt sind nur die Fälle, in denen in den Subgruppen fünf oder mehr Studierenden vertreten sind).

Elektro- und Informationstechnik

Gesamturteil nach Subgruppen



Erkennbar ist zunächst, dass zwischen beiden Subgruppen tatsächlich erhebliche Urteilsunterschiede bestehen. Allerdings zeigt die Graphik auch, dass beide Gruppen zu ähnlichen relativen Beurteilungen (verglichen mit den übrigen Fachbereichen) kommen, d.h. keiner der schlecht beurteilten Fachbereiche würde in die Spitzengruppe aufrücken, wenn man nur das Urteil der jüngeren Vollzeitstudierenden heranzieht. Ähnliches gilt für das Urteil der älteren Teilzeitstudierenden: Die Positionierung der Fachbereiche ändert sich nur geringfügig, wenn man ausschließlich das Urteil der älteren Studierenden berücksichtigt. Wie Abbildung 2 zeigt, ist trotz erheblicher Bewertungsunterschiede der Mittelwert aus allen studentischen Urteilen ein guter Schätzer für die Positionierung der Fachbereiche.

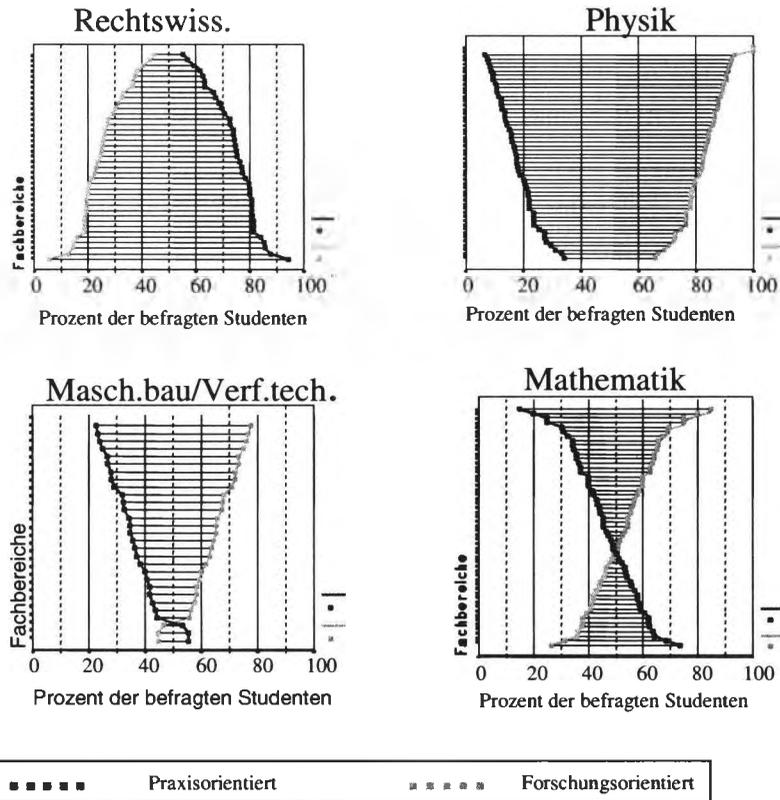
Die Existenz von Subgruppen mit unterschiedlichen Urteilsniveaus an sich ist daher kein Problem für eine grobe Positionierung der Fachbereiche. Dies ändert sich erst, wenn extrem inhomogene Fachkulturen vorliegen, d.h. wenn Subgruppen nicht nur völlig entgegengesetzt werten, sondern ihr Anteil an der jeweiligen Studierendenpopulation auch extrem schwankt.

Abb. 3: Wichtigkeit der Forschungs- und Praxisorientierung im Studium

Bewertungen der Studierenden verschiedener Disziplinen auf einer 6 stufigen Skala.

Praxisorientierung = % der Befragten, die den Praxisbezug für wichtig halten (Skalenwerte 1 & 2) und den Forschungsbezug für weniger wichtig halten oder Praxis- und Forschungsbezug für weniger wichtig halten.

Forschungsorientierung = % der Befragten, die den Forschungsbezug für wichtig halten (Skalenwerte 1 & 2) und den Praxisbezug für weniger wichtig halten oder Praxis- und Forschungsbezug für wichtig halten.



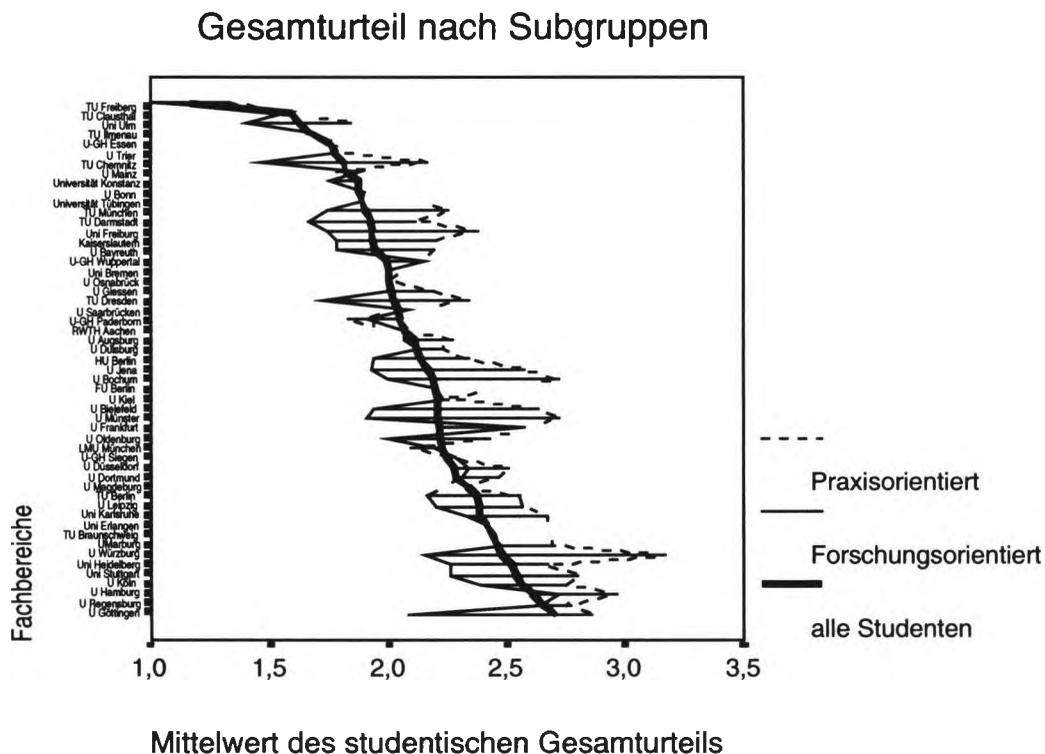
Derartige Phänomene lassen sich im Fach Mathematik beobachten: Zu den auffälligsten Einflussfaktoren in Tabelle 2 gehört die Einschätzung der Studierenden, ob eher der Forschungs- oder eher der Praxisbezug des Studiums wichtig sei. Von den Studierenden der Rechtswissenschaft halten 54% den Praxisbezug für wichtiger, von den Informatikstudierenden immerhin noch 28%. In der Physik hingegen halten 19% der Studierenden den Forschungsbezug für wichtiger, und die große Mehrheit stuft beide Aspekte gleichermaßen als sehr wichtig oder wichtig ein. Abweichend von diesem Muster finden sich in der Mathematik zwei Teilpopulationen von 16% und 24 % der Befragten, die entweder der Praxis oder der Forschung höheren Stellenwert einräumen. Diese Dichotomie reflektiert den enormen Bedeutungszuwachs, den die angewandte Mathematik neben der reinen Grundlagenmathematik erhalten hat. Während in der Rechtswissenschaft und in der Physik jeweils die Forschungs- bzw. Praxisorientierung klar dominiert, finden sich in der Mathematik sowohl Fachbereiche, in denen die Praxisorientierung

überwiegt, als auch Fachbereiche, in denen die Forschungsorientierung überwiegt⁵ (vgl. Abb. 3). Auch in der Mathematik vergeben die praxisorientierten Studierenden in der Regel schlechtere Noten. Vergleicht man nun die Urteilsmittelwerte von praxisorientierten Studierenden mit denen der forschungsorientierten an den einzelnen Fachbereichen, dann ergibt sich zwar immer noch eine signifikant positive Rangkorrelation zwischen beiden Gruppen (Kendalls Tau $b = .47$), aber die Zahl der Fachbereiche, die ihre Rangposition erheblich ändern, ist relativ groß: Rund 40% der vergleichbaren 38 Fachbereiche ändert die Rangposition um zehn oder mehr Plätze, wenn man die Positionierungen einmal aufgrund der Bewertungen der forschungsorientierten Studierenden durchführt und einmal auf Grund der Beurteilung der praxisorientierten Studierenden.

Die Zufriedenheit mit den Studienbedingungen hängt offenbar davon ab, ob es gelingt, den unterschiedlichen Erwartungshaltungen gerecht zu werden, die mit einer eher praktischen bzw. einer eher theoretisch-forschungsbezogenen Orientierung verbunden sind.

Abb. 4

Mathematik



Solche dramatischen Positionsveränderungen ergeben sich, weil die beiden Subgruppen sich nicht einfach durch eine Niveauverschiebung in ihrem Urteil auszeichnen, sondern durch eine fachbereichstypische Differenz in den Urteilen. An rund 40% der Fachbereiche liegen die Urteile beider Gruppen nur bis zu 0,3 Skaleneinheiten auseinander, aber an rund 24% der Fachberei-

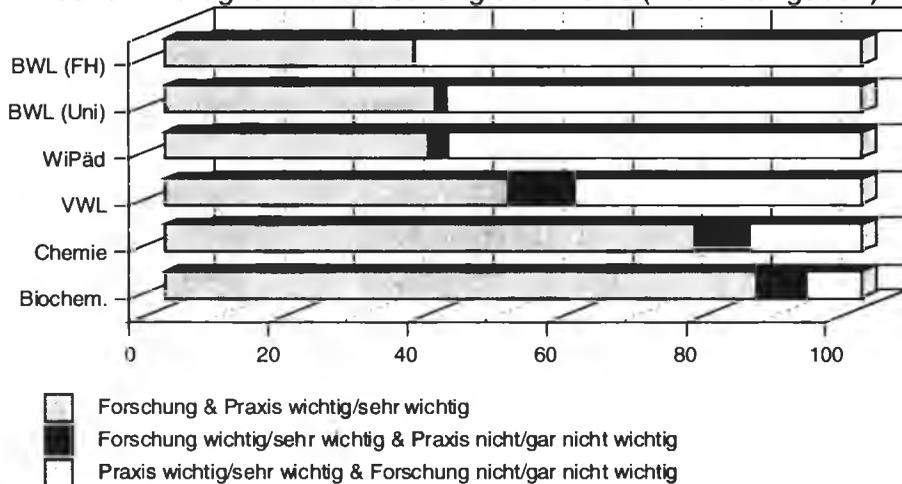
⁵ Dabei handelt es sich bei den hier wiedergegebenen CHE Untersuchungen ausschließlich um Studierenden mit dem Abschlussziel „Diplom“, die Berücksichtigung von Lehramtsstudierenden würde das Problem vermutlich noch mehr verschärfen.

che beträgt die Differenz 0,6 und mehr Skaleneinheiten (bezogen auf eine sechsstufige Antwortskala). Diese besonderen Verhältnisse in der Mathematik erklären auch, warum unterschiedliche Rankingstudien in diesem Fach zu unterschiedlichen Ergebnissen kommen.

Fazit: Die bisherigen empirischen Befunde deuten darauf hin, dass bias-Variablen einen allenfalls geringen Einfluss auf das Urteil der Studierenden haben und entsprechend bei einer Korrektur nur geringfügige Positionsveränderungen der Fachbereiche in Ranglisten zur Folge haben. Ausnahmen davon treten offenbar nur unter besonderen, fachspezifischen Bedingungen (vgl. Mathematik) auf. Ein Generalvorbehalt ist gegenüber der Aussagefähigkeit studentischer Beurteilungen ebenso ungerechtfertigt, wie gegenüber der Verwendung von Mittelwerten zur Charakterisierung der Fachbereiche.

Derartige Phänomene, wie sie in der Mathematik auftreten, werfen natürlich die Frage auf, ob die erfragten Sachverhalte aus der Sicht der Studierenden überhaupt relevant für die Qualität eines Studienangebotes sind. Folgt man den durch Marketingüberlegungen inspirierten Konzepten unterschiedlicher Rankingzielgruppen (vgl. Bayer in diesem Band), dann müsste jede Frage zusätzlich mit einer Relevanzeinschätzung verbunden werden. Items mit hoher Relevanz aus der Sicht der Studierenden erhalten dann ein größeres Gewicht als Items mit geringer Relevanz. Ein solches Vorgehen wurde im ersten Studienführer des CHE / Stiftung Warentest (1998) gewählt. Neben der Bewertung einzelner Sachverhalte wurden die Studierenden gebeten, auf einer vierstufigen Skala (sehr wichtig bis gar nicht wichtig) jedes Item mit einer Relevanzeinstufung zu versehen.

Abb. 5: Wichtigkeit von Forschung und Praxis (Prozentangaben)



Anm: Bewertung der Wichtigkeit des Forschungs- bzw. Praxisbezuges der Lehrveranstaltungen auf einer vierstufigen Skala von „sehr wichtig“ bis „gar nicht wichtig“. Quelle: CHE / Stiftung Warentest Studienbefragung 1998

Greift man zunächst einmal den oben schon dargestellten Bereich von Forschungs- und Praxisbezug der Lehrveranstaltungen heraus, dann zeigt sich wiederum eine erhebliche Differenz zwischen anwendungs- und forschungsorientierten Disziplinen: Während in der Betriebswirtschaft die Studierenden, an Fachhochschulen ebenso wie an den Universitäten, überwiegend den Pra-

xisbezug für wichtiger halten als den Forschungsbezug, werden in der Chemie und Biochemie überwiegend Praxis- und Forschungsbezug für gleichermaßen wichtig gehalten, zugleich finden sich zwei kleinere Teilpopulationen mit expliziter Forschungs- bzw. Praxisorientierung. Wenn diese Orientierungen zu unterschiedlichen Erwartungshaltungen an das Studium und in der Folge zu unterschiedlichen Bewertungen führen, so könnten sich auch die Mittelwerte der studentischen Beurteilungen auf Fachbereichsebene verändern, wenn man die unterschiedlichen Relevanzeinschätzungen der Studierenden berücksichtigt. Bei solchen Überlegungen muss allerdings sehr genau zwischen Effekten auf der Individualebene und solchen auf der Aggregatebene unterschieden werden. Auf der Aggregatebene wirkt sich eine Berücksichtigung der Relevanzeinschätzungen nur aus, wenn a) Subgruppen existieren, die sich in der Beurteilung und der Relevanzeinschätzung deutlich von ihren Kommilitonen unterscheiden, und b) wenn diese Subgruppen nicht in allen Fachbereichen mehr oder weniger gleichmäßig auftauchen, sondern sich an einzelnen Standorten konzentrieren. Nur unter diesen Bedingungen würde sich die Rangfolge der Fachbereiche tatsächlich ändern. Auf der Ebene von Gedankenexperimenten werden derartige Überlegungen gern als Beleg für die fehlende Aussagefähigkeit von Urteilsmitteiwerten angeführt.

Da beide Bedingungen aber in aller Regel jedoch nicht erfüllt sind, ergibt eine Berücksichtigung der studentischen Wichtigkeitseinschätzungen empirisch keine Veränderung der Positionierung der Fachbereiche. Dieser Befund ist unabhängig davon, welche Gewichtungsverfahren benutzt werden.

	Wichtigkeit als Gewichtungvariable	Multiplikative Gewichtung der Urteile	N
Inhaltliche Breite des Lehrangebotes	0,998 (p = 0,000)	0,902 (p = 0,000)	48
Forschungsbezug der Lehrveranstaltungen	0,993 (p = 0,000)	0,953 (p = 0,000)	48
Verfügbarkeit der benötigten Literatur	1,000 (p = 0,000)	0,964 (p = 0,000)	48
Ausstattung der Computerarbeitsplätze	0,998 (p = 0,000)	0,951 (p = 0,000)	48

Tab. 4: Korrelation (Pearson) zwischen gewichteten und ungewichteten Beurteilungen (Fachbereichsmittelwerte- BWL -Universitäten). Quelle: Studierendenbefragung CHE / Stiftung Warentest 1998

Tabelle 4 zeigt für verschiedene Fragekomplexe die Korrelationskoeffizienten zwischen den ungewichteten Mittelwerten der Fachbereiche und einer Gewichtung durch Mehrfachberücksichtigung der Fälle (Wichtigkeit als Gewichtungvariable) sowie einer multiplikativen Gewichtung der Beurteilung mit der Wichtigkeitseinschätzung. In beiden Fällen ergeben sich nahezu perfekte Korrelationsbeziehungen. Aufgrund dieser Befunde wurde in den folgenden Studienführern auf die Erhebung von Wichtigkeitsurteilen zu jedem Item verzichtet.

Fazit: Die Gewichtung von Urteilen mit den Relevanzeinschätzungen des beurteilten Sachverhaltes hat auf der Aggregatebene (Fachbereich) kaum Auswirkungen.

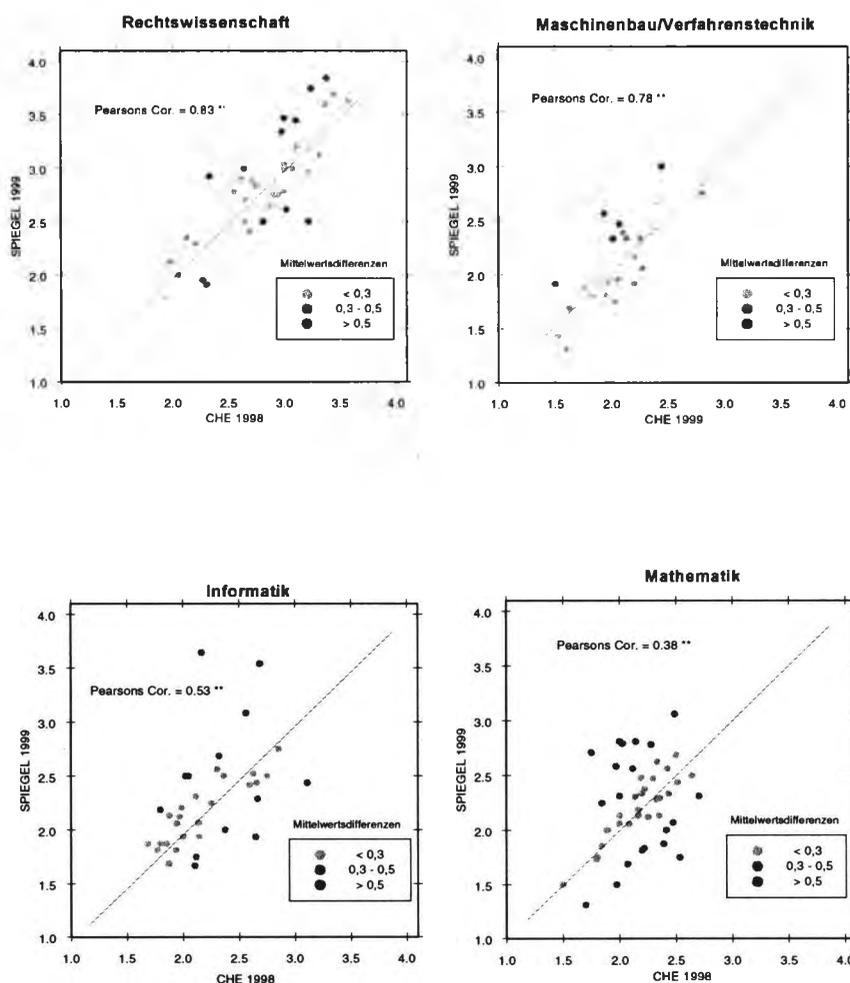
3. Einwand: Rankings führen zu instabilen, fast beliebigen Ergebnissen

Dieser Einwand beruht fast regelmäßig darauf, dass die an anderer Stelle eingeklagte Differenzierung von Rankingergebnissen beim Vergleich unterschiedlicher Studien völlig vernachlässigt wird. Ernsthaft wird wohl niemand erwarten, dass die Befragung unterschiedlicher Auskunftspersonen (von Managern über Professoren zu Studierenden), die Verfolgung unterschiedlicher Fragestellungen (von Imagestudien über indikatorengestützte Bewertungen bis hin zu subjektiven Einschätzungen) und unterschiedliche Verrechnungen von Einzelbefunden zu identischen Ergebnissen führen.

Vergleichen lassen sich nur Studien mit zumindest ähnlichem Design. Ein solcher Vergleich eröffnet zugleich die Möglichkeit die Reliabilität des Instrumentes bzw. den Einfluss verzerrter Stichproben zu prüfen. Die SPIEGEL Erhebung von 1999 und die CHE Studienführer 1999 und 2000 enthalten in der Studierendenbefragung ein ähnliches Item, in dem die Studierenden um eine Gesamtbeurteilung der Studiensituation gebeten werden. Ansonsten unterscheiden sich die Studien erheblich: Die SPIEGEL Befragung erfasst Studierenden aller Abschlussarten, die CHE Studie nur Diplom bzw. Staatsexamen; die SPIEGEL Studie basiert auf einer willkürlichen Auswahl der Studierenden mit Ortsvorgaben für die Interviewer, die CHE Studien beruhen Größenabhängig auf Vollerhebungen oder auf Zufallsstichproben, die SPIEGEL Umfrage ist als mündliche Campusbefragung durchgeführt worden, die CHE Studien als schriftliche, postalische Befragung, schließlich wurden die CHE Erhebungen mit einem weitaus größerem Stichprobenumfang durchgeführt und last not least: die Studien wurden zu unterschiedlichen Zeitpunkten durchgeführt, so dass Reaktanzen nicht ausgeschlossen werden können. Unterschiede also, die bei Methodenvergleichen in der Regel bereits deutliche Abweichungen in den Ergebnissen produzieren. Vergleicht man nun die Fachbereichsmittelwerte aus beiden Studien, dann zeigen sich für die Rechtswissenschaft und den Maschinenbau sehr hohe Übereinstimmungen (Pearson Cor. .78 bzw. .83; vgl. Abb. 6). Dieser Befund gilt auch für die Informatik mit der Einschränkung, dass hier zwei extreme Ausreißer den Korrelationskoeffizienten (.53) „drücken“. In der Mathematik hingegen ist zwar immer noch eine deutliche Korrelation zu erkennen, aber bei einer relativ großen Anzahl von Fachbereichen differieren die Fachbereichsmittelwerte um mehr als 0,5 Skalenpunkte. Die Ursachen für die starken Abweichungen in der Mathematik wurden oben bereits diskutiert, hinzu kommt, dass in der SPIEGEL Studie nicht nur Diplommathematiker sondern auch Lehramtsstudierende befragt wurden.

Fazit: Der Vergleich der beiden unabhängig durchgeführten und methodisch abweichenden Ranking Studien zeigt, dass Erhebungen mit vergleichbarer Fragestellung und einer zumindest ähnliche definierten Grundgesamtheit zu weitgehend übereinstimmenden Ergebnissen kommen.

Abb. 6: Vergleich der Studierendenbefragungen von CHE und SPIEGEL
Gesamtbeurteilung der Studiensituation
Mittelwerte und Mittelwertdifferenzen von Fachbereichen



4. Einwand: Die Stichproben sind zu klein oder verzerrt

Die vergleichsweise kleine absolute Anzahl der Befragten Studierenden gibt immer wieder Anlass über zu kleine oder verzerrte Stichproben und deren Auswirkungen zu spekulieren (vgl. Ott 1999). Das Design von Hochschulrankings hat es nicht mit einer, sondern mit sehr vielen Grundgesamtheiten (jeder Fachbereich bildet eine Grundgesamtheit) zu tun. Die Gesamtzahl der Befragten liegt daher regelmäßig weit über der Stichprobengröße, die etwa für repräsentative Bevölkerungsbefragungen üblich ist. Allein für die Fächer Elektrotechnik und Maschinenbau/Verfahrenstechnik wurden z.B. in der CHE Befragung rund 65.000 Studierende angeschrieben. Die Grundgesamtheit besteht allerdings nur aus den Studierenden im Hauptstudium. Deren Zahl liegt z.B. im Fach Maschinenbau an mehr als der Hälfte der Standorte unter 300 Studierenden, an mehr als 20% der Standorte sogar unter 100 Studierenden. An Standorten mit

weniger als 300 Studierenden im Hauptstudium wurden für den CHE Studienführer Vollerhebungen durchgeführt, an den übrigen Stichproben von 300 Studierenden gezogen. Das sind Stichprobengrößen, die die Standards der Umfrageforschung bei weitem übererfüllen.

Bis 50 Studierende	13,0 %
51 bis 100 Studierende	8,9 %
101 bis 200 Studierende	18,7 %
201 bis 300 Studierende	17,9 %
Über 300 Studierende	41,5 %

Tab. 5: Fachbereiche nach Zahl der Studierenden zwischen dem 5. und 14. Fachsemester im Fach Maschinenbau, Universitäten und Fachhochschulen 1998. Quelle: Statistisches Bundesamt

Ein Problem stellt allerdings der niedrige Rücklauf in den Studierendenbefragung dar (Studierenden je nach Fach 20% bis 25% Rücklauf; Professoren 40% bis 50%). Postalische Befragungen weisen grundsätzlich niedrigere Ausschöpfungsraten auf als persönlich-mündliche Befragungen. Empirisch variieren die Ausschöpfungsquoten postalischer Befragung zwischen 10% und 90%, und man geht von durchschnittlichen Werten in der Größenordnung von 47% aus (Porst 1999). Daran gemessen ist der Rücklauf in der Professorenbefragung befriedigend, in den Studierendenbefragungen jedoch nicht. Allerdings folgt aus einer niedrigen Ausschöpfungsraten keineswegs zwangsläufig auch eine Verzerrung der Stichprobe. Empirisch lässt sich zeigen, dass der sog. „Nonresponse-Bias“ keineswegs mit der Höhe der Ausschöpfungsquote abnimmt (Koch 1998). „Mehr“ bedeutet daher keineswegs auch „besser“. Wie der Vergleich mit der SPIEGEL Befragung (s.o.) gezeigt hat, ergeben sich zunächst keine Hinweise darauf, dass durch systematische Selbstselektion nur ein bestimmter Studierendentypus an den Befragungen teilgenommen hat. Die Behauptung, dass Stichproben allein aufgrund der niedrigen Ausschöpfungsraten verzerrt seien, ist eine Vermutung ohne Beleg. Replikationsstudien und der Vergleich unterschiedlicher Stichprobenziehungsverfahren sprechen bisher nicht für eine Verzerrung.

Fazit: Der Stichprobenumfang des CHE Studienführers ist mehr als ausreichend. Die niedrigen Ausschöpfungsraten sind unbefriedigend, wenn auch nicht ungewöhnlich. Anhaltspunkte für eine systematische Verzerrung liegen bisher nicht vor.

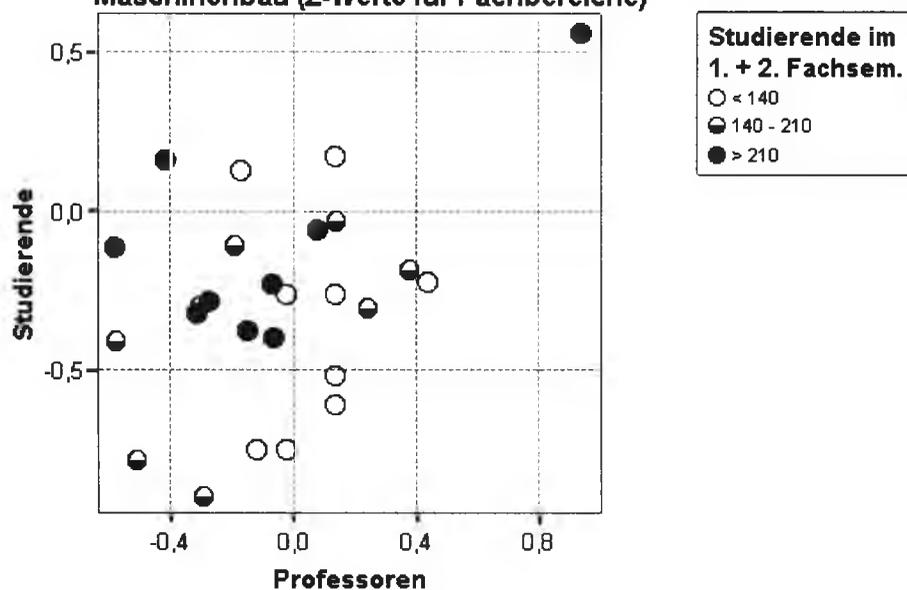
5. Einwand: Man braucht gar kein Ranking, die Größe der Hochschule bestimmt die Urteile der Studierenden

Hochschulrankings auf der Basis von Studierendenurteilen kommen immer wieder zu dem Ergebnis, dass kleinere Fachbereiche von den Studierenden tendenziell besser bewertet werden als große. Zur Erklärung wird darauf verwiesen, dass den Studierenden an den großen Fakultäten die Nestwärme fehle und sie allenfalls die Rahmenbedingungen, die nur indirekt etwas mit Qualität zu tun haben, beurteilen würden (vgl. Meinefeld 2000).

Betrachtet man einmal zwei Studiengänge genauer, einen stark nachgefragten mit hohen Studierendenzahlen pro Fakultät (Rechtswissenschaft) und einen weniger stark nachgefragten mit

vergleichsweise niedrigen Studierendenzahlen (Maschinenbau), dann lässt sich zeigen, dass solche Erklärungen wenig empirischen Rückhalt haben: Zunächst einmal zeigen die Graphiken 7 und 8, dass die Beziehung zwischen Größe und Bewertung keineswegs in allen Fächern gilt, sondern vor allen Dingen in den großen, überlaufenen Studienfächern. Im Maschinenbau sind derartige Beziehungen kaum zu erkennen. Sodann deuten die gleichgerichteten Urteile von Professoren und Studierenden darauf hin, dass auch die Professoren an großen Universitäten die Lehrsituation schlechter beurteilen. Dies kann nicht auf fehlende Nestwärme zurückzuführen sein. Schließlich zeigt sich sowohl im Maschinenbau wie auch in der Rechtswissenschaft, dass es großen Hochschulen durchaus gelingen kann, positive Urteile von Studierenden und Professoren zu erhalten. Die unterschiedlichen Bewertungen der großen Hochschulen zeigen zudem, dass auch unter widrigen Bedingungen sich sehr unterschiedliche Ergebnisse produzieren lassen. Schlussendlich spricht etwa in der Rechtswissenschaft auch nichts dafür, dass große Fakultäten eine bessere, von den Studierenden nicht gewürdigte, Lehrleistung erbringen; die Noten im Staatsexamen unterscheiden sich nicht signifikant von denen an kleineren Fachbereichen.

Abb. 7: Gesamturteile zur Lehre von Studenten und Professoren im Maschinenbau (Z-Werte für Fachbereiche)



CHE Studienführer 2000

Fazit: Die Größe von Fachbereichen determiniert keineswegs die Urteile von Studierenden und Professoren, vielmehr scheint sich nur in den stark frequentierten Studienfächern die Größe der Fachbereiche auf die Lehrqualität auszuwirken, und selbst dann sind große Fachbereiche durchaus in der Lage, Lehr- und Betreuungsangebote zu entwickeln, die von Studierenden und Professoren positiv beurteilt werden.

Gesamturteile zur Lehre von Studenten und Professoren in der Rechtswissenschaft (standardisierte Z-Werte für Fachbereiche)

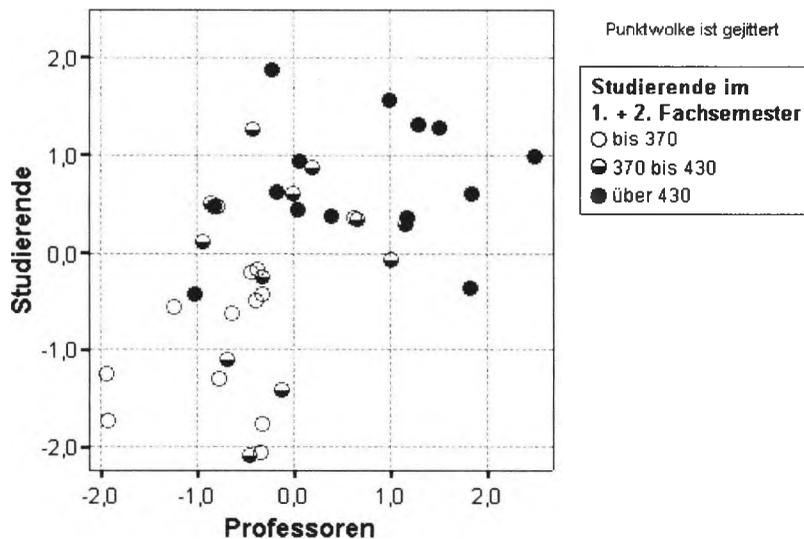


Abb. 8: Rechtswissenschaft

6. Einwand: Mittelwert- und Indexbildungen nivellieren die Urteile

Die Fülle von Informationen, die in Rankingstudien typischerweise erhoben wird, muss für den Nutzer in geeigneter Form verdichtet werden. Das gilt sowohl für die Zahl der Variablen als auch für die Charakterisierung der studentischen Urteile an einem Fachbereich.

Erstere Informationskomprimierung wird üblicherweise durch die Entwicklung von Indices oder Skalen bewerkstelligt, die möglichst eindimensional die Eigenschaften des Bewertungsobjektes wiedergeben sollen, letztere über die Berechnung von Mittelwerten und geeigneten Angaben zur Homogenität bzw. Streuung der gemittelten Werte. Indizes, wie sie auch im CHE Studienführer benutzt werden, durchlaufen also eine zweimalige Mittelwertbildung. Die Folge ist, dass nicht mehr erkennbar ist, ob mittlere Urteile oder sehr heterogene Urteile dem Mittelwert zugrunde liegen. Zwar informieren die berechneten Konfidenzintervalle darüber, wie einheitlich die Indexwerte ausgefallen sind (s.u.), aber nicht mehr darüber, ob die Befragten typische Urteilsprofile aufweisen (vgl. auch Kromrey 2000).

Es ist deshalb sicherlich wünschenswert nach Möglichkeiten einer informationshaltigeren Aufbereitung der Daten zu suchen. Kromrey (vgl. den Beitrag in diesem Band) schlägt dazu vor, die Urteilenden aufgrund von clusteranalytisch gewonnenen Urteilstypologien in verschiedene Beurteilungskluster einzuordnen und eine Bewertung der Fachbereiche auf die quantitativ dominanten Cluster zu stützen. Für eine Routineberichterstattung ist dies Verfahren leider nicht tauglich. Clusteranalysen reagieren sehr empfindlich auf die gewählten Fusionierungsverfahren, Ähnlichkeitsmaße und die Behandlung fehlender Werte. Hinzu kommt, dass ein solches Verfahren – je nach Fach – bis zu 60 Einzelurteile der Studierenden zu einem Profil verarbeiten müsste. Unter solchen Bedingungen entstehen entweder beliebige oder triviale Ergebnisse, denn einfache Antwortmuster wie „alles gut“ oder „alles schlecht“ gibt es so gut wie nicht. Ihre Existenz

würde auch die Frage nahe legen, ob hier nicht ein „response set“ wirksam ist, also eine stereotype Bewertung unabhängig von den gestellten Fragen. In aller Regel antworten die Studierenden sehr differenziert. Selbst wenn man die Kriterien für „alles gut“ sehr weit fasst, bleibt diese Gruppe in der Minderheit. Betrachtet man beispielsweise nur einmal die Fragenkomplexe zur Lehre, Studienorganisation, Raumausstattung und Bibliothek, dann antworten von den ca. 23.000 Studierenden der CHE Befragung 1999 nur knapp 20% so, dass mindestens zwei Drittel der Items mit den Skalenwerten 1 und 2 benotet werden, während ca. 80% gute und schlechte Noten so vergeben, dass keine durchgängige Präferenz für eine Note erkennbar ist. Angesichts der oben genannten Fallzahlen produzieren Clusteranalysen schließlich für kleine Fachbereiche ein zusätzliches Problem: Die Fallzahl in den einzelnen Clustern kann sehr niedrig werden, eine Interpretation ist dann kaum mehr möglich.

Am Beispiel des Index Lehre aus der Studierendenbefragung im Maschinenbau lassen sich Nutzen und Problematik clusteranalytischer Gruppierung demonstrieren⁶. Tabelle 6 zeigt eine Vier-Clusterlösung für die acht Einzelitems, die im Index Lehre verrechnet werden⁷. Cluster 4 weist eine sehr positive Urteilstendenz auf, das Gegenstück (Cluster 2) zeigt hingegen eine sehr kritische Beurteilung der acht Items. Cluster 1 und Cluster 3 zeigen die erwartete Differenzierung in den mittleren Urteilen. Im Cluster 1 wird etwas schlechter als im Durchschnitt geurteilt, wobei insbesondere die Praxisdimension (Lehrveranstaltungen durch Praktiker, Praxisbezug, Interdisziplinarität) sehr kritisch beurteilt wird. Im Cluster 3 hingegen wird etwas besser als im Durchschnitt geurteilt (mit Ausnahme der Didaktik und des Angebots an Projektseminaren). In der Zeile „Index“ sind zum Vergleich die Mittelwerte des Index Lehre angegeben; sie fassen die Urteilstendenz in diesen Clustern durchaus treffend zusammen. Insgesamt stellen die Cluster also eine Mischung aus Niveaushiftung der Urteile und schwach ausgeprägten Urteilsprofilen dar.

	Maschinenbau Cluster			
	Cluster 1 25% der Personen	Cluster 2 11,5% der Personen	Cluster 3 43 % der Personen	Cluster 4 21 % der Personen
	Mittelwert	Mittelwert	Mittelwert	Mittelwert
Lehrangebot Projektseminare etc.	2,87	4,17	2,55	1,63
Lehrangebot inhaltliche Breite	2,07	2,66	1,89	1,51
Lehrangebot Forschung	3,19	3,82	2,49	1,81
Lehrangebot Praxisbezug	3,61	4,71	2,64	1,77
Lehrangebot internat. Ausrichtung	3,30	4,20	2,89	1,83
Lehrangebot Interdisziplinarität	3,14	4,25	2,44	1,76
Lehrangebot Vermittlung Lehrstoff	3,47	4,48	2,71	2,07
Lehrangebot Praktiker-Veranstalt.	4,07	4,76	2,79	1,75
INDEX	3,22	4,13	2,55	1,77

Tab. 6 (in der Zeile Index sind die Mittelwerte für den berechneten Index Lehre angegeben)

⁶ Das Beispiel von Kromrey (in diesem Band) kombiniert den Index Lehre mit dem Index Studienorganisation, so dass die unterschiedlichen Ranggruppen nicht vergleichbar sind. Der Index Lehre, der im CHE Studienführer benutzt wurde, besteht aus acht einzelnen Items. Für die folgenden Analysen wurden nur Fälle berücksichtigt, die auf allen acht Items valide Werte aufweisen.

⁷ Vgl. dazu den Methodenbericht zum CHE Studienführer: CHE Arbeitspapier Nr. 22, 2000

Betrachtet man nun, wie sich die Befragten an den einzelnen Hochschulen auf die Cluster verteilen (vgl. Tabelle 7), dann zeigt sich, dass erwartungsgemäß an den gut beurteilten Hochschulen (kleiner Indexmittelwert) das Cluster 4 überwiegt und umgekehrt an den schlecht beurteilten die Cluster 2 und 1. Allerdings ist Cluster 2 an keiner Hochschule stärker besetzt als die übrigen Cluster; das würde bedeuten, dass eine Schlussgruppe gar nicht existiert. Eine solche Interpretation wird den Daten aber erkennbar nicht gerecht, denn an den schlecht bewerteten Hochschulen (z.B. Hannover, Berlin, Wuppertal) findet sich eine deutliche Mehrheit in den kritisch urteilenden Clustern 1 und 2, während typische Mittelfeld Universitäten (z.B. Duisburg, Dresden) eine deutliche Mehrheit im Cluster 3 aufweisen.

Tab 7: Maschinenbau - Index Lehre
Prozentuale Besetzung der Cluster, Indexmittelwerte und Fallzahlen

			% in Cluster 1	% in Cluster 2	% in Cluster 3	% in Cluster 4	N ^a
Index	2,13	TU Clausthal	8,3	2,3	39,4	50,0	132
Lehre	2,14	Bergakad. Freiberg TU	11,4	,0	38,6	50,0	44
MW	2,20	U Kiel	20,0	,0	40,0	40,0	5
	2,30	TU Ilmenau	25,9	,0	33,3	40,7	27
	2,36	U Bremen	20,0	2,9	45,7	31,4	35
	2,42	U Rostock	17,4	8,7	34,8	39,1	23
	2,46	TU Chemnitz	17,6	2,9	52,9	26,5	34
	2,60	U Saarbrücken	23,1	3,8	50,0	23,1	26
	2,62	UGH Siegen	22,7	6,8	43,2	27,3	44
	2,62	TU Hamburg-Harburg	22,7	7,3	50,9	19,1	110
	2,63	RWTH Aachen	20,4	14,3	40,8	24,5	49
	2,63	BTU Cottbus	25,6	6,1	46,3	22,0	82
	2,66	UGH Paderborn	11,5	15,4	42,3	30,8	26
	2,69	TU Dresden	19,8	7,2	61,3	11,7	111
	2,70	UGH Duisburg	22,0	10,0	50,0	18,0	50
	2,73	U Kaiserslautern	25,7	5,7	48,6	20,0	35
	2,77	TU Darmstadt	33,3	10,4	37,5	18,8	48
	2,82	U Dortmund	25,6	14,1	34,6	25,6	78
	2,84	UGH Kassel	19,4	13,9	52,8	13,9	36
	2,86	U Erlangen-Nürnberg	28,9	10,0	44,4	16,7	90
	2,87	U Stuttgart	26,5	17,1	46,2	10,3	117
	2,90	TU Braunschweig	25,9	19,0	36,2	19,0	58
	2,92	U Karlsruhe (TH)	28,3	18,9	37,7	15,1	106
	2,93	U Bochum	30,0	20,0	35,0	15,0	20
	2,97	UGH Essen	29,0	16,1	45,2	9,7	31
	3,01	TU München	41,5	11,3	37,7	9,4	53
	3,18	U Hannover	26,9	26,9	34,6	11,5	26
	3,27	TU Berlin	41,6	23,0	31,9	3,5	113
	3,36	UGH Wuppertal	31,3	25,0	40,6	3,1	32

^a. Nur Fälle, die auf allen 8 Variablen, die für die Indexbildung benutzt wurden, valide Werte aufweisen.

Für einen Vergleich liegt es daher nahe, die Cluster 1 und 2 zu einer Schlussgruppe zusammenzufassen. Gruppiert man nun die Hochschulen danach, welches Cluster den größten Anteil der Befragten auf sich vereinigen kann, dann ergibt sich eine Spitzengruppe (Cluster 4 dominiert), eine Mittelgruppe (Cluster 3 dominiert) und eine Schlussgruppe (die Summe von Cluster 1 und 2 dominiert). Diese Gruppeneinteilung lässt sich mit der im CHE Studienführer angewandten Ranggruppenbildung auf der Grundlage von Konfidenzintervallen um den Mittelwert (s.u.) vergleichen.

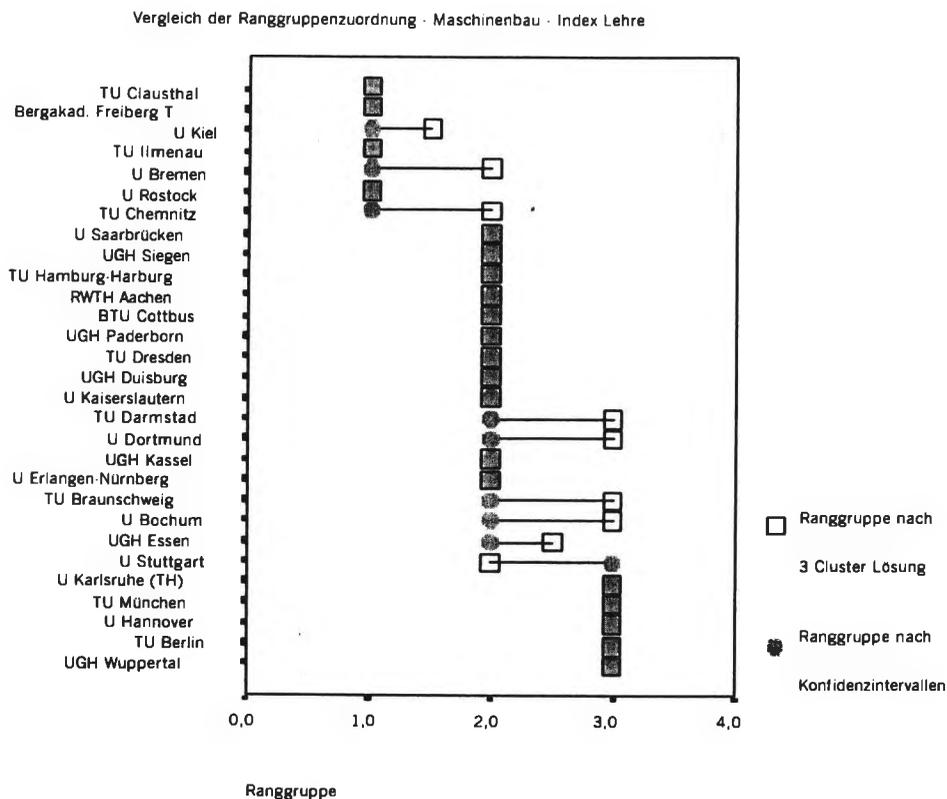


Abb. 9: Ranggruppen nach Clustermodell und nach Konfidenzintervallen

Das Ergebnis ist in Abbildung 9 wiedergeben. Erkennbar ist zunächst, dass die clusterbasierte Gruppenzuordnung keine grundsätzlich anderen Ergebnisse hervorbringt, als die auf Mittelwerten und Konfidenzintervallen basierende Gruppeneinordnung. Wechsel finden nur zwischen der Mittelgruppe und den Extremgruppen statt, nicht aber zwischen den Extremgruppen. Insgesamt überwiegt die Zahl gleich klassifizierter Hochschulen deutlich die Zahl der Abweichungen. Die Abweichungen treten bevorzugt an den Gruppengrenzen auf, also dort, wo eine Abgrenzung ohnehin schwerfällt. Weiterhin ist erkennbar, dass von den neun Hochschulen, die die Ranggruppe wechseln, zwei nicht eindeutig verortet werden können (in Kiel und Essen weisen zwei Cluster die gleiche Besetzungshäufigkeit auf). Und schließlich ist zu berücksichtigen, dass die Hochschulen Essen, Bochum, Chemnitz, Bremen und Kiel⁸ in dieser Reanalyse (vgl. die Ausschlusskriterien und Tab. 7) nur noch mit Stichprobengrößen von ≤ 35 Personen repräsentiert sind. Die clusterbasierte Gruppenzuordnung wird damit sehr instabil: Beispielsweise kann an der Universität Bochum eine geringfügige Veränderung der Urteile von nur zwei Personen dazu führen, dass ein anderes Cluster dominant wird. Auch an der Universität Stuttgart liegt das Kritikercluster (43,6%) und das mittlere Cluster (46,2%) extrem nah beieinander. Ähnliches gilt für Dortmund und Darmstadt.

⁸ Die Universität Kiel bietet keinen Studiengang Maschinenbau an. Sie ist in diesem Sample nur enthalten, weil unter die Definition „Studienbereich Maschinenbau“, wie sie vom CHE festgelegt wurde, auch der Studiengang „Materialwissenschaft“ gehört, der in Kiel angeboten wird.

Ein Zugewinn an Informationen entsteht also nicht auf der Ebene von Ranggruppenzuordnungen, sondern nur dann, wenn man die Besetzung aller vier Cluster im Auge behält. Dann lässt sich etwa die überwiegend durchschnittlich beurteilte TU Dresden von den sehr heterogen beurteilten Hochschulen Braunschweig oder Darmstadt unterscheiden. Dieser Informationszugewinn bedeutet aber auch, dass aus einer Kennzahl (Index Lehre) vier Kennzahlen (Besetzungshäufigkeiten in den vier Clustern) werden. Dieser Weg der Darstellung bietet sich also für vertiefende Information an, allerdings stellt sich die Frage, ob dann nicht an Stelle der tendenziell instabilen und von vielen Konstruktionsentscheidungen abhängigen Clusteranalyse eine einfache Darstellung der Notenverteilung innerhalb eines Indexwertes nicht ebenso informativ, wenn nicht sogar informativer ist. Abbildung 10 zeigt eine solche Verteilung für einige der oben analysierten Daten. Vergleicht man die nahe beieinander im Mittelfeld platzierten Hochschulen Aachen und Dresden, zeigt sich sehr deutlich, dass Aachen einen weitaus höheren Anteil an „Einser-Noten“ auf den acht Items des Index Lehre verbuchen konnte als Dresden.

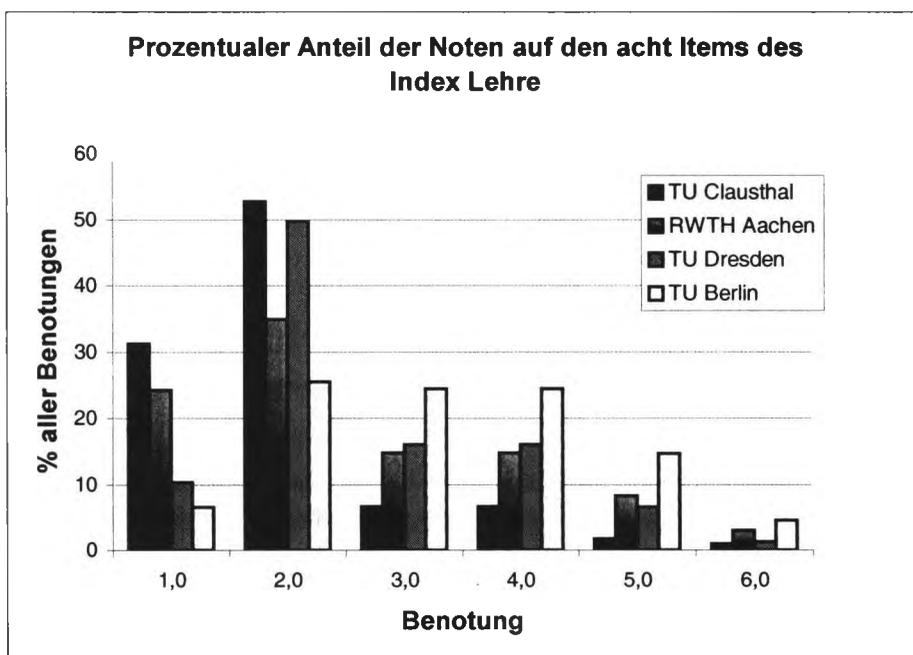


Abb. 10

Ein anderer Weg, die Informationen über die Urteilsstreuung besser zu nutzen, besteht darin, theoretisch plausible und empirische vorfindbare Differenzierungen für eine Subgruppenbildung zu nutzen (vgl. oben das Beispiel Mathematik). Man überlässt dann die Gruppenbildung nicht den Unwägbarkeiten der Clusteranalyse, sondern benutzt Grundorientierungen (z.B Praxis-versus Forschungsorientierung), um homogene Subpopulationen zu erzeugen.

Fazit: Index- und Mittelwertbildungen stellen Informationsverdichtungen dar, die notwendigerweise auch zu Informationsreduktionen führen. Der Vergleich von clusteranalytischen Verfahren und varianzanalytisch basierten Verfahren zeigt aber, dass bei einer hohen Informationsverdichtung – wie sie für orientierende Überblicke notwendig ist – keine gravierenden Differenzen auftreten. Cluster- und Subgruppenanalysen stellen sehr wünschenswerte ergänzende und vertiefende Information zur Verfügung, einen Ersatz für

Mittelwertbildungen stellen sie jedoch aufgrund der Verfahrensabhängigkeit und Instabilität nicht dar.

7. Einwand: Die Hochschulen unterscheiden sich kaum, Ranggruppen führen zu arbiträren Urteilen

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. Deshalb werden Konfidenzintervalle um die Fachbereichs-Mittelwerte der jeweiligen Urteile ermittelt, und auf dieser Grundlage wird 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. Die Größe des Konfidenzintervalls informiert zudem darüber, ob hier weitgehend mittlere Urteile abgegeben wurden oder ob sehr heterogene Urteile zu einem mittleren Mittelwert geführt haben. 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, entweder mit Hilfe einer CD Rom, die dem Studienführer des CHE beigelegt ist oder über ein „persönliches Ranking“ auf der Internetseite des Studienführers (<http://www.stern.de/servlet/CHE2>).

Auch Statistikern fällt es jedoch schwer, dieser Interpretation zu folgen. So fragt z.B. Jensen (vgl. den Beitrag in diesem Band), warum in der Elektrotechnik die Universitäten Karlsruhe und Siegen mehr mit der Universität Chemnitz gemeinsam haben sollten, als mit Magdeburg oder Kiel. Ein Blick auf die Konfidenzintervalle beantwortet diese Frage sehr schnell: Eine solche Behauptung wurde nie aufgestellt. Die Konfidenzintervalle zeigen sehr klar, dass Karlsruhe und Siegen **weniger** mit der Universität Chemnitz gemeinsam haben als mit Magdeburg oder Kiel. Ausgesagt ist lediglich, dass Karlsruhe und Siegen besser bewertet sind als die Schlussgruppe (z.B. Hannover, Wuppertal), was sich eben von Kiel nicht behaupten lässt.

Diese Befunde sind darüber hinaus weitgehend verfahrensunabhängig. Berechnet man anstelle der Konfidenzintervalle von Verteilungsannahmen unabhängige Bootstrap-Rangkonfidenzintervalle (vgl. Jensen in diesem Band), entstehen extrem ähnliche Ergebnisse (so sticht beispielsweise hier wie dort die BTU Cottbus mit einem sehr breiten Konfidenzintervall heraus).

Elektrotechnik - Universitäten

95% Konfidenzintervall um den Mittelwert des stud. Gesamturteils

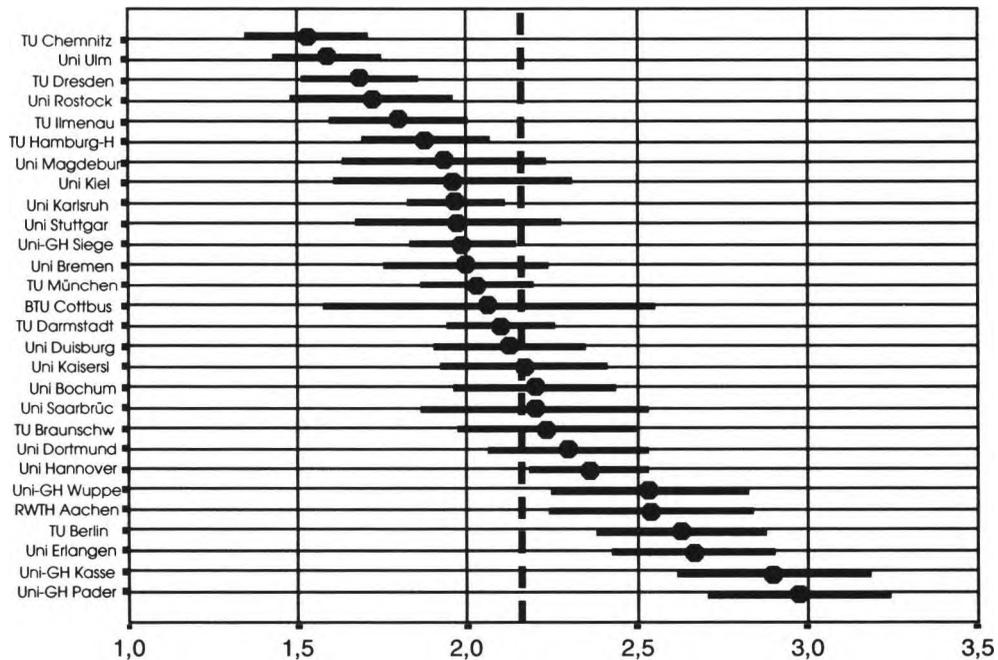


Abb. 11

Diese Ergebnisse sind auch durchaus plausibel, denn anders als in der BWL oder Rechtswissenschaft liegen in den Ingenieurwissenschaften vergleichsweise homogene Verhältnisse vor. Ein Ranking hat nun nicht die Aufgabe, künstliche Differenzen aufzubauen, sondern durchaus auch die Aufgabe, homogene Verhältnisse abzubilden. 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 Bedürfnis Hochschulen auf ein Siegereppchen zu stellen, das aufgrund von Rangplätzen besetzt wird (vgl. Meinefeld): Wenn die publizierten Hitlisten – wie im Falle des CHE-Studienführers – keine Rangplätze enthalten, dann werden sie einfach durch Auszählung im Nachhinein ermittelt, um dann 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 (vgl. ebd.).

Was jedoch unbefriedigend an jeder Gruppeneinteilung bleibt, das sind die überkonturierten Grenzziehungen zwischen den Gruppen. Dieses Problem entsteht bei jeder Gruppenbildung, unabhängig vom gewählten Verfahren: So entscheiden z.B. im von Kromrey vorgeschlagenen Clustermodell extrem wenige Personen darüber, ob ein Cluster die 40% oder 50% Marke übersteigt. Zur nachfolgenden Hochschule besteht dann i.d.R. natürlich kein signifikanter Unter-

schied. Ranggruppen stellen lediglich eine Dichtomisierung zwischen sehr gut und sehr schlecht bewertet her. Dabei bleiben die Grenzen zur Mittelgruppe notwendigerweise unscharf. Um einen Vergleich zwischen einzelnen Hochschulen zu ermöglichen, wäre es daher sinnvoll neben den Ranggruppen auch die Konfidenzintervalle für die einzelnen Mittelwerte zu präsentieren (vgl. z.B. Hornbostel 1998, 1999b; Müller-Böling und Hornbostel 2000). Dann würde sehr schnell deutlich, dass Universitäten wie Kiel oder Magdeburg aufgrund der heterogenen Urteile und der entsprechend breiten Konfidenzintervalle in der Mittelgruppe positioniert wurden.

Fazit: Rankingkritiker scheinen sich nur schwer damit abfinden zu können, dass zwischen sehr vielen Hochschulen keine signifikanten Bewertungsunterschiede bestehen. Die üblicherweise berechneten Konfidenzintervalle geben jedoch – auch im Vergleich mit anderen Gütebeurteilungen – sehr zuverlässig an, wo Unterschiede behauptet werden können und wo nicht. Wünschenswert wäre es, wenn diese Information – ergänzend zu den Ranggruppen - auch in den publizierten Rankings zu finden wäre.

8. Einwand: Zu viel subjektive Urteile, zu wenig harte Fakten.

In der Tat enthält der Studienführer des CHE überwiegend subjektive Urteile. Zwar werden eine ganze Reihe von statistischen Rahmendaten ausgewiesen, ebenso die Fachstudiedauer, aber andere Kernbereiche werden nicht mit Fakten versehen. Das hat verschiedene Gründe: Manche Informationen wie etwa die Teilnehmerzahlen an Veranstaltungen oder Universitäts-Eingangstests, wie in den USA üblich, existieren einfach nicht, andere Informationen wie etwa Abbruchquoten sind aufgrund gesetzlicher Regelung (Einstellung der Individualstatistik) nicht verfügbar, wieder andere existieren nur rudimentär, wenngleich sie sehr wünschenswert wären wie z.B. Informationen über den beruflichen Werdegang der Absolventen.

Es gibt auch eine Reihe von Informationen, die prinzipiell zugänglich sind, aber nicht zu interpretationsfähigen Indikatoren verarbeitet werden können. Dazu gehört beispielsweise, dass es bisher nicht gelungen ist, eine Betreuungsrelation, die auch die Dienstleistungsverflechtungen angemessen berücksichtigt, einheitlich zu definieren und zu erfassen (also das tatsächliche numerische Verhältnis von Professoren und zu betreuenden Studierenden, nicht die normativ fixierten Curricularnormwerte).

Strukturdaten sind ein Beispiel dafür, dass es nicht sinnvoll ist, von vornherein die Interpretation solcher Kennzahlen auf einer „gut-schlecht“ Dimension festzulegen, sondern dem Informationsnutzer Interpretationsspielräume zu überlassen. Ein hoher Frauenanteil unter den Studierenden ist zunächst weder gut noch schlecht, er kann aber eine solche Konnotation bekommen, wenn etwa eine Schülerin sich für ein männerdominiertes Fach interessiert und bei der Standortentscheidung den Frauenanteil berücksichtigt. Ähnliches gilt für die Größe von Fachbereichen und Hochschulen, die Studiendauer etc.

Werden Kennzahlen jedoch für evaluative Zwecke eingesetzt, ist eine solche Beliebigkeit äußerst unerwünscht. Und dies ist denn auch die Achillesferse „harter Daten“. Beispielsweise wurde versucht für den CHE Studienführer die Qualität der Bibliothek zusätzlich zu den subjektiven Einschätzungen durch harte Fakten aus der Bibliotheksstatistik zu ergänzen. Nur leider ist es nur begrenzt möglich – abgesehen davon, dass Institutsbibliotheken gar nicht erfasst werden – Abonnementszahlen, laufende Ausgaben oder Monographiebestände als Qualitätsindika-

toren zu interpretieren. Harte Fakten sind nicht per se besser als subjektive Einschätzungen, wenn es darum geht die Eignung eines Angebotes für die Bedürfnisse der Nutzer zu beurteilen.

Was allerdings an „harten Fakten“ verfügbar und interpretierbar, steht auf der CD Rom des Studienführers auch zur Verfügung.

9. Einwand: Rankings reduzieren das komplexe Leistungsspektrum von Hochschulen auf studentische Beurteilungen.

Rankingkritiker stellen häufig auf die Mehrdimensionalität hochschulischer Leistungen ab, die mit studentischen Befragungen allein nicht eingefangen werden kann. Das ist zweifellos richtig, erstaunlicherweise wird dabei allerdings die wichtigste Leistungsdimension neben der Lehre, nämlich die Forschung, kaum thematisiert. Auch US-amerikanische Rankings berücksichtigen diese Dimension in der Regel nicht. Dass Forschung und Lehre ein Kuppelprodukt bilden, heißt weder, dass sich automatisch Humdoldtsche Ideale einer die Lehre befruchtenden Forschung einstellen, noch umgekehrt, dass sich Forschung und Lehre in einer Art antagonistischer Zuspitzung gegenseitig behindern (vgl. Schimank 1995, Güdler/Sack1996). Vielmehr existiert eine bunte Mischung aus allen denkbaren Kombinationen.

Forschungsleistungen stellen nicht nur die wichtigste Quelle für das Renommee einer Hochschule dar, sie sind auch für jene Studierenden mit einer starken Forschungsorientierung, ebenso wie für Doktoranden, Hochschulwechsler und last not least für die an Forschung interessierten Dritten eine wichtige Information (vgl. Weingart 1995, Hornbostel 1999c).

Forschungsleistungen lassen sich insofern einfacher als Lehrleistungen erfassen, als diese i.d.R. öffentlich zugänglich sind (publiziert) und eine überlokale community laufend über die Qualität – zumindest eines Teils - dieser Forschungsergebnisse, Urteile fällt (Manuskriptbegutachtungen, Drittmittelbewilligungen, Tagungseinladungen, Auszeichnungen, Zitierungen etc.). Allerdings gilt für Forschungsindikatoren ebenso wie für die Lehre, dass sie a) fachspezifisch justiert werden müssen (bzw. nur fachspezifisch einsetzbar sind), b) geeignete (meist umstrittene) Bezugsgrößen definiert werden müssen (Personal), c) fast immer subdisziplinäre Differenzierungen vorliegen, die ähnlich wie bei den Lehrindikatoren die Frage nach geeigneten Maßzahlen bzw. Differenzierungen in der Darstellung aufwerfen.

Das zeigt sich beispielsweise deutlich bei einer Analyse der Drittmiteleinwerbungen. Sie sind vor allen Dingen deshalb interessant, weil sie anders als etwa bibliometrische Indikatoren, die nur retrospektiv berichten, sehr zeitnahe oder sogar prospektive Informationen liefern. Bei diesem Indikator ist allerdings strittig, ob man ihn als Input oder Output Messung interpretieren kann. Für erstere Deutung spricht, dass die Miteleinwerbungen noch nichts über den Ertrag eines Projektes aussagen. Für letztere spricht, dass erstens der Drittmittelbewilligung i.d.R. ein aufwändiger Begutachtungsprozess zugrunde liegt, zweitens mit dem Drittmittelantrag häufig bereits ein sehr voraussetzungsvolles Zwischenprodukt vorliegt. Ein Indiz dafür, dass die Begutachtung von Drittmittelanträgen durchaus prognostische Validität besitzt, ergibt sich, wenn man einmal den Publikationsoutput daraufhin untersucht, ob die aus drittmittelgeförderten Projekten hervorgegangenen Publikationen eine andere fachliche Resonanz erzeugen als die übrigen Publikationen (vgl. Hornbostel 1997).

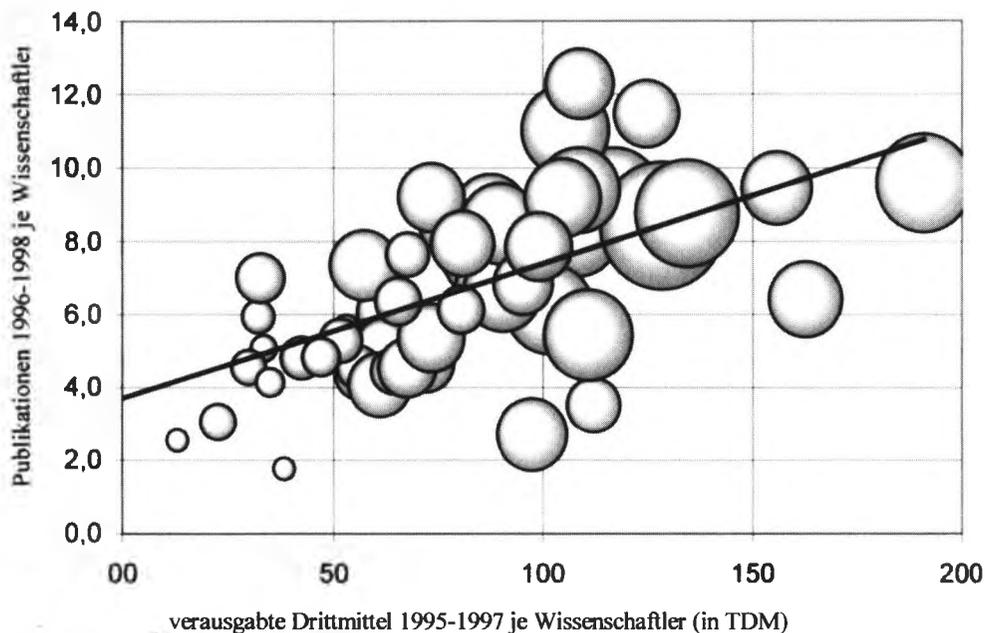
In welchem Umfang aus Drittmittelprojekten Publikationen in hochrangigen Zeitschriften resultieren, hängt allerdings auch davon ab, wie die Publikationsgepflogenheiten im jeweiligen Spezialgebiet beschaffen sind, welche Drittmittelanteile für apparative Zwecke und sonstige

Infrastruktur benötigt werden, dem time-lag zwischen Forschungsprojekt und Publikationszeitraum und schließlich natürlich vom Erfolg des Forschungsprojektes. Abb. 12 zeigt für die Physik die Anzahl der Beiträge in Zeitschriften (pro Professor am Fachbereich), die im Science Citation Index ausgewertet werden und der Höhe der verausgabten Drittmittel. Deutlich erkennbar ist die Beziehung zwischen Drittmittelvolumen und Publikationen, aber auch der Umstand, dass im mittleren Bereich pro „Drittmittelmark“ sehr unterschiedlich hohe Publikationszahlen realisiert werden.

Gewichtiger als subdisziplinäre Differenzen sind die Disziplinunterschiede. Abb. 13 zeigt die Antragsaktivität in den einzelnen Disziplinen. Die Daten basieren auf einer Vollerhebung unter den Professoren der jeweiligen Fachgebiete im Rahmen des Studienführers des CHE (vgl. Hornbostel 2001a). Von den rechtswissenschaftlichen Professoren haben mehr als 60% in den letzten drei Jahren keine Drittmittelanträge gestellt. Dieser Anteil liegt bei den Maschinenbauern unter 10%, dafür haben aber fast 40% der Maschinenbauprofessoren mehr als 10 Drittmittelanträge innerhalb von drei Jahren gestellt. Dieser unterschiedlichen Antragsintensität korrespondieren natürlich entsprechende Unterschiede im Gesamtvolumen der eingeworbenen Drittmittel.

Von den befragten Professoren aus den Bereichen Maschinenbau/Verfahrenstechnik und Elektrotechnik machen nur 3,3% keine Angaben bzw. haben kein Forschungsprojekt in den Jahren 1997 bis 1999 durchgeführt. Eine Forschung ohne Drittmiteleinsatz findet in den Ingenieurwissenschaften also praktisch nicht statt.

Abb. 12: Physik: Fachbereiche nach Drittmitteln und Publikationen
(Blasenfläche = Drittmittel absolut)

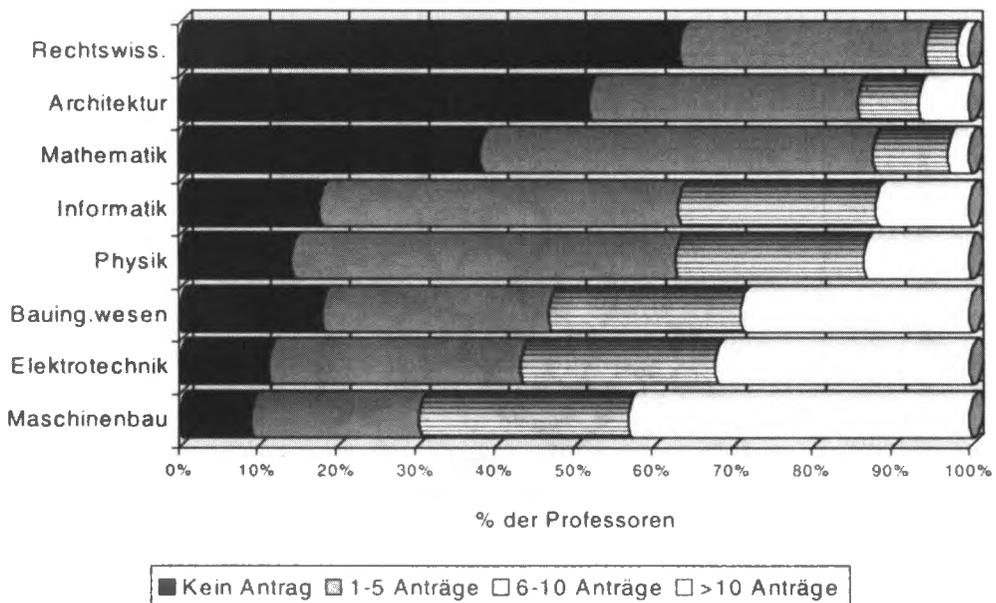


Voraussetzung für die Nutzung eines Drittmittelindikators ist, dass Drittmittelforschung in den untersuchten Fachgebieten üblich ist und nicht ausschließlich im Rahmen subdisziplinärer Spezialisierung (insbes. empirische Forschung) betrieben wird. Wie der Abbildung 13 zu entnehmen ist, sind die Rechtswissenschaften und die Architektur fast drittmittelabstinent. Aus diesem Grund macht es z.B. keinen Sinn, die Drittmittelinwerbungen in der Rechtswissenschaft als Indikator für Forschungsleistungen zu benutzen.

Auch die Nutzung der verschiedenen Finanzierungsquellen fällt disziplinspezifisch aus. Wie Abbildung 13 zeigt, haben die kaum Drittmittelforschung betreibenden Juristen keine Schwerpunkte bei der Wahl ihrer Finanziers. Physiker und Mathematiker hingegen stellen ihre Drittmittelanträge vornehmlich bei der DFG. Die Ingenieurwissenschaften schließlich akquirieren ihre Mittel überwiegend in der Privatwirtschaft (vgl. dazu den Beitrag von Güdler in diesem Band). Da diesen Mitteln sehr unterschiedliche Antragsbegutachtungen zugrunde liegen, stellt sich auch die Frage, ob allen Drittmittelzuflüssen eine ähnliche Indikatorqualität für Forschungsleistungen attestiert werden kann (vgl. Hornbostel 2001b).

Ähnlich wie für die Drittmittelinwerbungen gilt für die Publikationen, die man zunächst einmal als Aktivitätsindikator im Forschungsbereich interpretieren kann, dass disziplinspezifische Besonderheiten zu berücksichtigen sind: Wie die Abbildung 14 zeigt, ist in der Physik mit Artikeln in Fachzeitschriften und Kongressbeiträgen der ganz überwiegende Teil der Publikationen erfasst. Zudem ist die Reputation der verschiedenen journals nicht sonderlich umstritten, so dass z.B. mit dem Science Citation Index oder spezialisierten Fachdatenbanken relativ problemlos „wichtige“ Beiträge identifiziert werden können.

Abb. 13 Drittmittelanträge von Universitätsprofessoren in den letzten 3 Jahren
(Prozent der befragten Professoren)



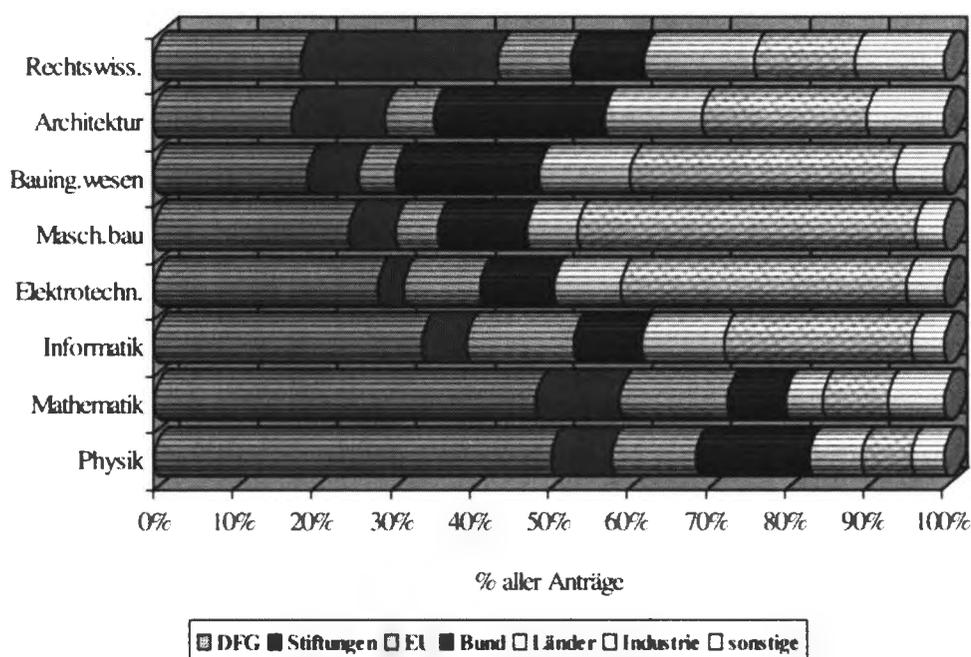
Quelle: Hornbostel, CHE Studienführer 1999 und 2000
Erhebungszeitraum: Masch.bau, Elektrotechnik, Architektur 1999, alle anderen 1998

In den Ingenieurwissenschaften spielt sich ein ganz erheblicher Teil der wissenschaftlichen Kommunikation in Kongressbänden ab. Ganz anders die Situation bei den Rechtswissenschaft-

lern: Dort spielen Festschriften (Sammelbände) und Monographien und als Reflex darauf Rezensionen eine wichtige Rolle. Artikel in Fachzeitschriften streuen über die unterschiedlichsten journals (auch nicht-juristische). Die Unterscheidung zwischen originärer Forschung, Kommentierung und Dokumentation ist schwer zu ziehen. Über die Wertigkeit einzelner Zeitschriften herrscht wenig Einverständnis. Wollte man die juristischen Publikationen zu einem Indikator verarbeiten, wäre es nicht nur notwendig, die unterschiedlichen Publikationstypen auf irgendeine Weise zu verrechnen, sondern auch den diversen Zeitschriften Gewichte zuzuordnen. Im CHE Studienführer werden daher jeweils fachspezifische Publikationsindikatoren gebildet und auch nur dann publiziert, wenn diese Indikatoren aussagekräftig sind.

Abb. 14

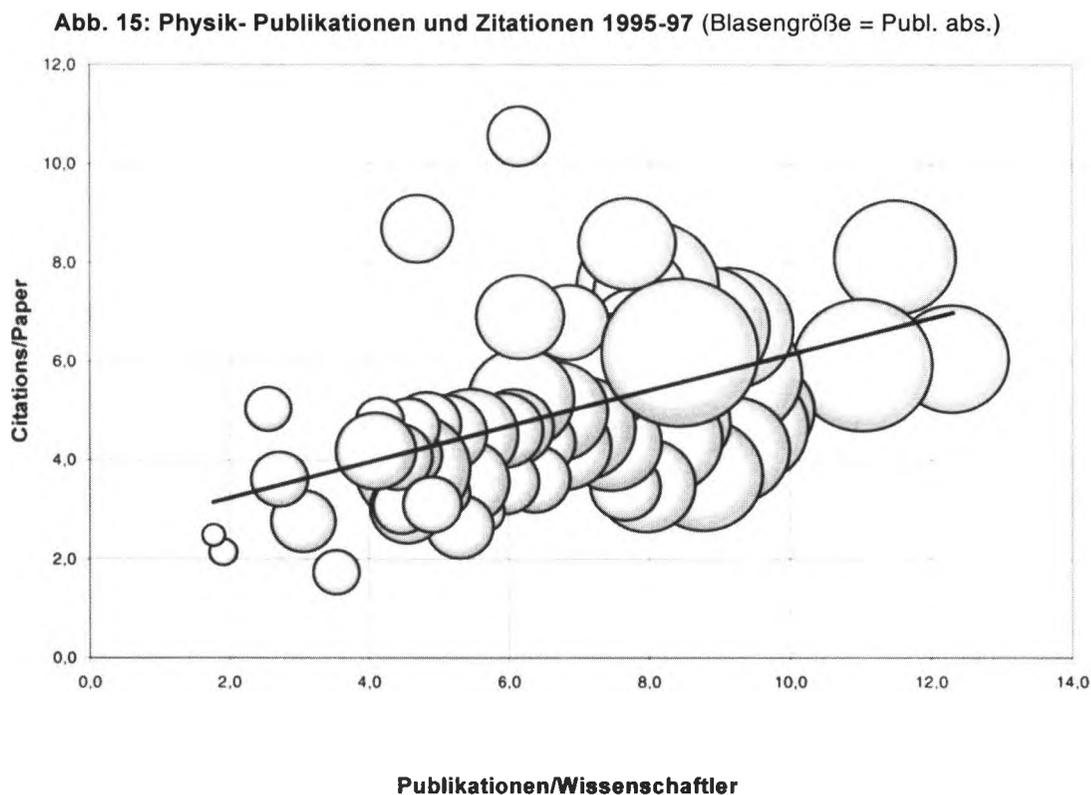
Drittmittelanträge von Universitätsprofessoren nach Drittmittelgebern



Quelle: Hornbostel, CHE Studienführer 1999 und 2000. Erhebungszeitraum: Masch.bau, Elektrotech., Bauing., Architektur 1999, alle anderen 1998

Publikationszählungen wie im Studienführer des CHE (Physik), die auf dem Science Citation Index beruhen, enthalten bereits eine Qualitätskomponente, denn nur weltweit stark zitierte Zeitschriften werden in diese Datenbank aufgenommen. Allerdings erzeugen auch in vielzitierten Zeitschriften die einzelnen Beiträge eine sehr unterschiedliche Resonanz (Zitationen). Für eine Berichterstattung über Forschungsleistungen wird man auch diese Qualitätseinschätzung berücksichtigen müssen. Möglich ist das jedoch nur dort, wo auch ein hinreichender Anteil der Publikationen im Science Citation Index erfasst ist, das heißt überwiegend in den Naturwissenschaften. Abbildung 15 zeigt anhand der Daten des CHE Studienführers, dass zwischen Publikationsintensität und der durchschnittlichen Resonanz (erhaltene Zitate) der Artikel zwar eine positive Beziehung besteht, auch hier zeigen sich aber gerade im mittleren Bereich ganz erhebliche Unterschiede in der erreichten Resonanz in der Fachöffentlichkeit.

In Fachgebieten, für die Zitationsanalysen nicht möglich oder nicht sinnvoll sind, kann man derartige Qualitätsindikatoren entweder gar nicht nutzen oder nur auf Umwegen – über Gewichtungungsverfahren – Annäherungen versuchen.

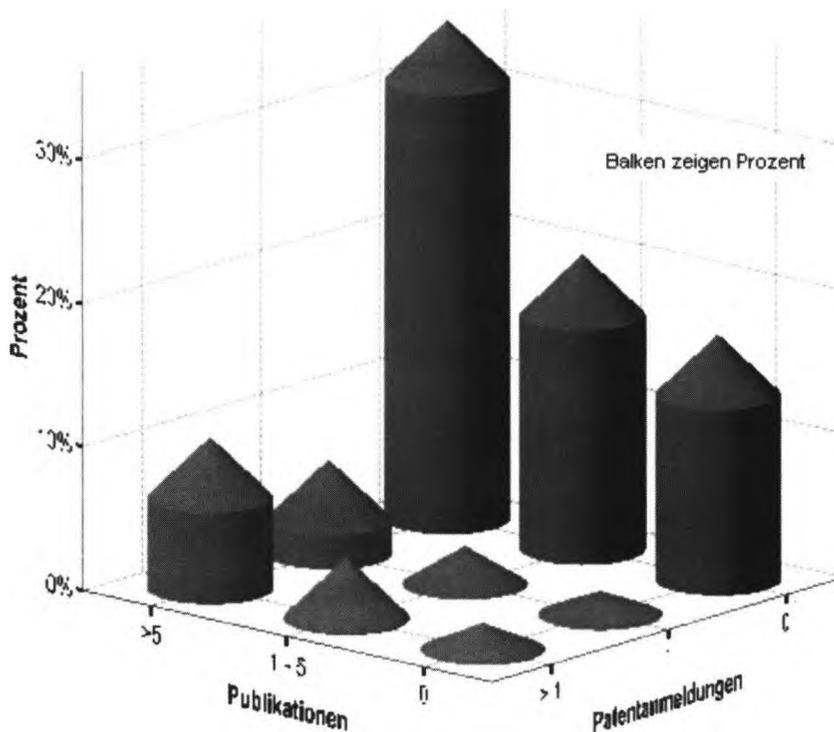


In stark anwendungsorientierten Disziplinen stellt sich ein zusätzliches Problem: Ein Teil der Forschungsarbeiten fließt nicht primär in das wissenschaftliche Kommunikationssystem, sondern in Anwendungskontexte, häufig im Rahmen von Kooperationen mit Wirtschaftsunternehmen. Einen Anhaltspunkt über das Ausmaß dieser innovativen, anwendungsorientierten Forschung lässt sich über die Zahl der Patentanmeldungen der Professoren gewinnen (vgl. dazu den Beitrag von Schmoch in diesem Band). Ein eigenständiger Indikator ergibt sich insbesondere deshalb, weil Publikationen patentschädlich sein können, insofern also ein Spannungsverhältnis besteht hinsichtlich der Sicherung von Prioritätsansprüchen. Anwendbar ist ein solcher Indikator aber nur in Disziplinen, in denen Patentanmeldungen regelmäßig und in einem erheblichen Umfang vorkommen. Das gilt etwa für den Maschinenbau, die Elektrotechnik, die Medizin und auch für die Chemie (vgl. Becher u.a. 1996).

Die Gewinnung entsprechender Daten ist ausgesprochen mühsam, da in den einschlägigen Datenbanken nicht ohne weiteres erkennbar ist, ob eine Patentanmeldung auf einen Hochschulprofessor zurückgeht oder nicht. Für die Patentanalyse in den Fächern Elektrotechnik und Maschinenbau des Studienführers des CHE wurde daher eine namentliche Überprüfung aller Hochschulprofessoren in der Patentdatenbank des Deutschen Patentamtes (PATDPA) durchgeführt. Ausgewählt wurden alle veröffentlichten deutschen Patent- und Gebrauchsmusteranmeldungen sowie alle veröffentlichten europäischen und PCT-Patentanmeldungen (Patent Cooperation Treaty der World Intellectual Property Organization) mit Priorität (bzw. prioritätsentsprechender Anmeldung) in den Jahren 1995 bis 1997. Dabei wurden mehrere Veröffentlichungen zu einer Priorität sowie eine Veröffentlichung zu mehreren Prioritäten jeweils nur einmal gezählt.

Abb. 16

**Patentanmeldungen und Publikationen in Fachzeitschriften 1996-98
Universitätsprofessoren aus dem Fachgebiet Maschinenbau/Verfahrenstechnik**

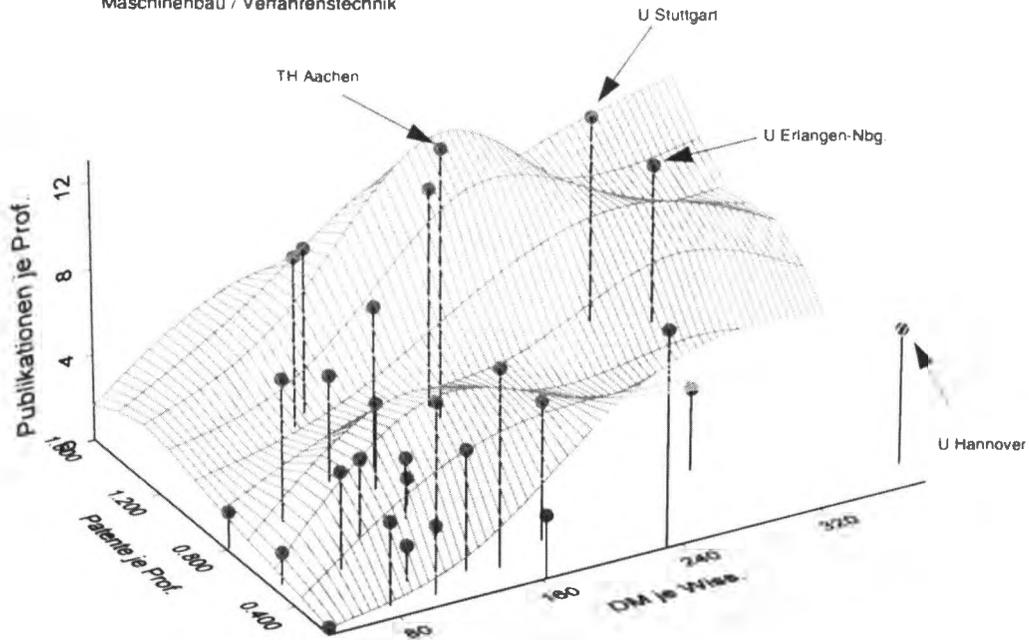


Im Ergebnis zeigt sich, dass nur ca. 28% der Professoren (Maschinenbau/Verfahrenstechnik an Universitäten) im Untersuchungszeitraum eine oder mehrere Patentanmeldungen eingereicht haben. Von diesen gehört allerdings nur eine Minderheit (ca. 13%) zu jenem Typus, der ausschließlich Patentanmeldungen vornimmt und nicht mit Publikationen auch im wissenschaftlichen Kommunikationssystem präsent ist. Wie Abbildung 16 zeigt, ist der größte Teil der patentaktiven Professoren auch mit Publikationen vertreten. Allerdings lassen sich für immerhin 17% der Maschinenbauprofessoren (nur Universitäten) weder Patentanmeldungen noch Publikationen im Untersuchungszeitraum nachweisen.

Angesichts dieser Fülle von Einzelinformationen stellt sich – ähnlich wie für die Lehrindikatoren – die Frage, wie man einerseits den fachlichen und innerfachlichen Differenzierungen mit aussagekräftigen Indikatoren gerecht werden kann und andererseits auch für den Nicht-Spezialisten einen orientierenden Überblick über die Forschungsleistungen bieten kann. Abbildung 17 vereinigt drei ausgewählte Indikatoren (Publikationen, Patente und Drittmittel). Erkennbar ist, dass nur sehr wenige Hochschulen auf ausschließlich einem Indikator hohe Werte erreichen, in der Regel harmonisieren die verschiedenen Messungen durchaus. D.h. die einzelnen Messungen unterschiedlicher Aspekte der Forschungsleistung gehen zwar nicht in einander auf, korrelieren aber deutlich untereinander. Diese Übereinstimmung lässt sich für einen weniger differenzierten, orientierenden Überblick nutzen.

Abb. 17

Drittmittel je Wissenschaftler, Patentanmeldungen je Prof., Publikationen je Prof.
 Maschinenbau / Verfahrenstechnik



Zugleich erlaubt eine solche Zusammenfassung die Minimierung von Messfehlern einzelner Indikatoren, die kaum vermeidbar sind. Sie ist im übrigen auch kompatibel mit den Reputationsurteilen, die die Professoren über jeweils anderen Hochschulen abgeben (vgl. oben, Tab. 2). Der Unterschied besteht darin, dass der Hochschullehrtipp nur jeweils eine sehr kleine Gruppe von Hochschulen erfasst, während die Forschungsindikatoren zwar Leistungsunterschiede in der gleichen Richtung signalisieren, aber erstens mit weniger starkem Gefälle zwischen Hochschulen und zweitens auch mit markanten Abweichungen zwischen der Bekanntheit und dem Leistungspotential. Abb 18 zeigt diese Zusammenhänge für die im CHE Studienführer untersuchten ingenieurwissenschaftlichen Studiengänge. Besonders stark empfohlene Hochschulen (wie etwa TH Stuttgart oder TH Aachen) weisen auch hohe Werte für Publikationen und Patentanmeldungen pro Professor auf, andere Hochschulen (wie z.B. UGH Paderborn, Maschinenbau oder Universität Ulm, Elektrotechnik) weisen zwar ebenfalls hohe Publikationszahlen und Patentanmeldungen pro Professor auf, werden aber von relativ wenigen Professoren als Studienort empfohlen. Reputation, wie sie sich in den Studienortempfehlungen der Professoren ausdrückt, verändert sich nur langsam.

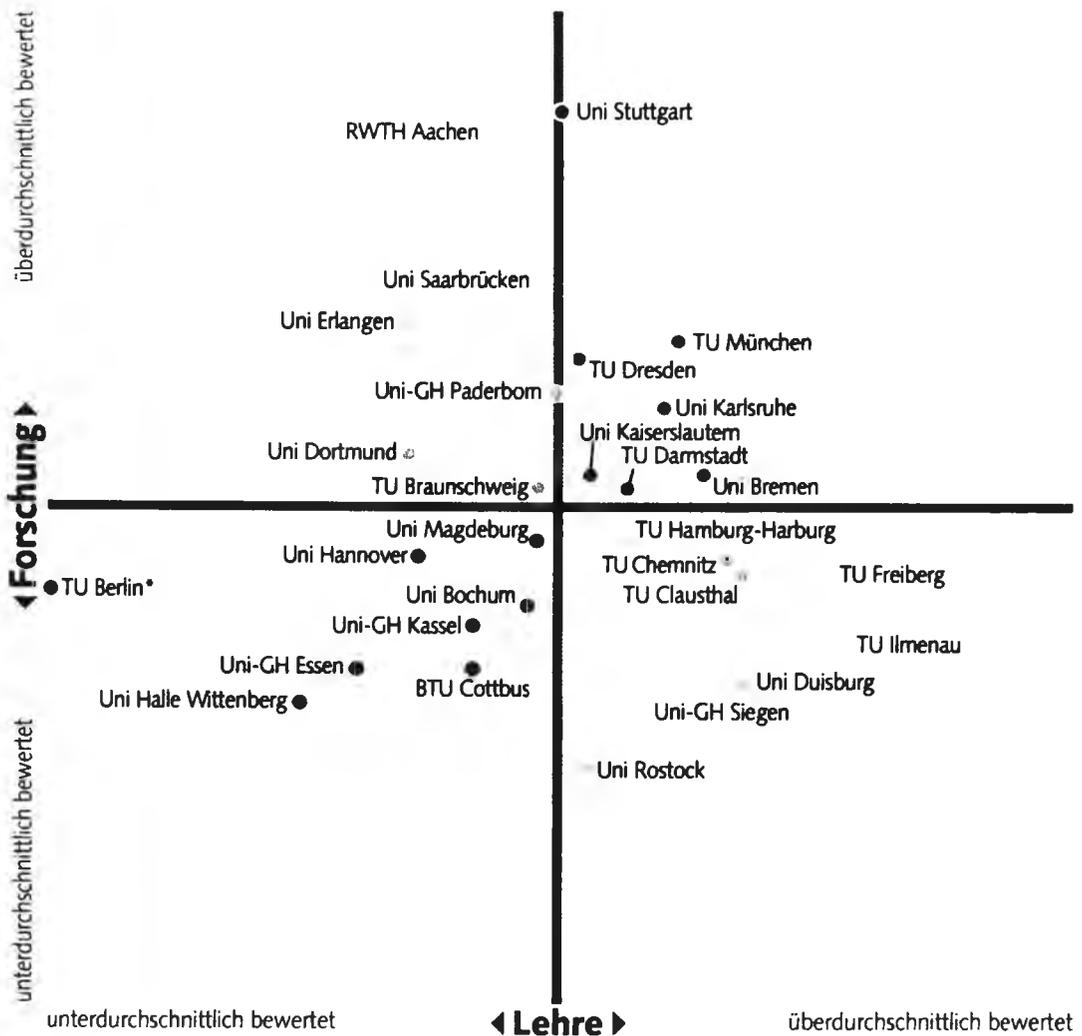
10. Einwand: Rankings fehlt die Zielgruppenorientierung

Das Grundmodell für diese Forderung ist im Beitrag von Bayer in diesem Band skizziert. Es geht von einer Marketingstrategie aus und bemüht dazu Analogien aus dem Wirtschaftsbereich. Abiturienten oder Studierenden sind wie die Käufer von Autoreifen mit bestimmten Präferenzen ausgestattet (von sportlich bis sparsam) und sehen sich einem differenzierten Marktangebot gegenüber. Die Bewertung der Produkte hängt vollständig an den Präferenzen der Käufer. D.h. der gleiche Autoreifen wird von den einen als sehr gut von den anderen als sehr schlecht be-

wertet. Aus diesem Modell wird nun einerseits hergeleitet, dass Rankings nicht aussagekräftig seien, weil auch dann, wenn der Prozentsatz von sparsamen Käufern und sportlichen gleich groß ist, keine entscheidungsunterstützende Information entstünde, andererseits wird daraus die Forderung abgeleitet, adressatenspezifische Kriterienkataloge zu entwickeln. Das erste Argument ist schlicht und logisch falsch: Würden die sportlichen Käufer die Billigreifen immer schlecht bewerten, dann käme bei einer Mittelwertbildung die Urteilsdifferenzierung der sparsamen Käufer zum tragen, würden die sparsamen alle hochwertigen und teuren Reifen schlecht bewerten kämen die Urteilsdifferenzierung der Sportwagenfahrer zum tragen. Bei gegebenen Präferenzen wäre genau die entscheidungsrelevant Information entstanden, die gesucht wurde. Schlicht ist das Argument, weil eine Studienentscheidung nun einmal unter etwas komplexeren Bedingungen verläuft als ein Reifenwechsel. Die empirisch wichtigste Differenzierung auf der Käuferseite ist die Unterscheidung von akademischen und praktischen Präferenzen. Diese Differenz ist aber schon in den Institutionen abgebildet (Fachhochschule, Universität). Um im Bild zu bleiben: Nicht die Urteile aller Reifenkäufer werden gemittelt, sondern getrennt, die der Kunden spezialisierter Reifenhändler. Abgesehen davon, dass sich ein Kunde gelegentlich verirrt, sind damit die wichtigsten Präferenzunterschiede bereits berücksichtigt, denn niemand hat bisher Universitäts- und Fachhochschulranglisten durcheinandergewürfelt. Die verbleibenden Bewertungsunterschiede bei den Studierenden folgen konsequenterweise auch nicht ausgeprägten Urteilsprofilen, sondern stellen überwiegend Niveaushiftungen dar, so dass in den meistens Fächern eine Aufbereitung der Daten für einzelne Gruppen (vgl. die Clusteranalyse) die Zahl der Indikatoren explosionsartig wachsen lässt, den Informationsgehalt (Rangliste) jedoch kaum verändert. Tod durch Informationsentropie wäre die Diagnose für ein solches Ranking angesichts des Bedarfs an einfachen Orientierungsgrößen, die Abiturienten in Gruppendiskussionen über den CHE-Studienführer einfordern. Auch aus einem weiteren Grund ist die von Bayer geforderte Zielgruppenorientierung wenig erstrebenswert. Es geht beim Ranking nämlich nicht um Marketing – auch wenn viele Hochschulen die Ergebnisse für diese Zwecke nutzen –, sondern um die Bereitstellung entscheidungsrelevanter Informationen. Was entscheidungsrelevant ist, sollte man aber mündigen Konsumenten durchaus selbst überlassen. Die Degradierung zur „Zielgruppe“, die dann einen Indikatorencocktail bekommt, den die Evaluateure für angemessen halten, ist aus einer Marketingperspektive erstrebenswert, aber nicht aus der Perspektive von Information und Orientierung suchenden Interessenten. Der CHE Studienführer in der Online-Version überlässt daher den Informationssuchenden die Kombination von Kriterien für eine Auswahl, darunter auch Kriterien, die nichts mit der Qualität der Ausbildung zu tun haben, sondern mit sozialen oder finanziellen Belangen.

Die Forderung nach immer weiteren Differenzierung beruht auf einem Missverständnis der Funktion von Rankings. Diese ersetzen nämlich keineswegs die Evaluation an den Hochschulen oder die Studienberatung. In hochschulinternen Evaluationen kann eine detaillierte Analyse der Zusammenhänge zwischen Erwartungen und Voraussetzungen der Studierenden und deren Beurteilungen zu steuerungsrelevantem Wissen führen (vgl. Krempkow 1999). Ortsvergleichende Evaluationen hingegen müssen einen Spagat bewältigen zwischen einer möglichst hohen Komprimierung von Informationen und einer Differenzierung und Detaillierung für Informationssuchende, die Vertiefung bereits vorhandenen Wissens suchen.

Faktorenanalyse Maschinenbau



* Der Faktorwert für den Faktor Lehre liegt bei der TU Berlin außerhalb des Darstellungsbereichs

Abb. 18

Rotierte Komponentenmatrix *

	Faktor	
	Forschung	Lehre
Publikationen je Professor **	.830	
Studienortempfehlung Professor	.739	
Drittmittel je Professor	.591	
Patentanmeldungen je Professor	.589	
Studierendenurteil Lehre		.825
Professorenurteil Lehre	-.112	.810

* Hauptkomponentenanalyse, Varimax Rotation mit Kaiser-Normalisierung ** mit Autorenzahl gewichtet

Fehlende Werte durch Mittelwerte ersetzt, Erklärter Varianzanteil: Faktor Forschung 33%, Faktor Lehre 22%

Ein Beispiel für eine Informationsverdichtung, die auch auf Ranggruppenbildung verzichtet, sind die im CHE Studienführer 2000 benutzten Faktorenanalysen über die Hauptdimensionen Forschung und Lehre (vgl. Abb.18). Für diese Faktoranalysen werden sowohl subjektive Urteile von Professoren und Studierenden als auch „harte Fakten“ (Forschungsindikatoren) verrechnet. Übrig bleiben die „Konsenszonen“ verschiedener Beurteilungen. Damit ist ein Ausgangspunkt gegeben für eine Informationserschließung, die zumindest in einem elektronischem Medium auch sehr differenzierte und detaillierte Datenpräsentationen erlaubt. Anstelle von vordefinierten Zielgruppen entsteht auf diese Weise ein orientierendes Informationssystem, das Fakten ebenso wie die subjektiven Urteile unterschiedlicher Bewertergruppen (und wo geboten auch von Subpopulationen) bereitstellt.

Literatur:

- Becher, G., Gering, T., Lang, O. / Schmoch, U. Patenwesen an Hochschulen. Hrsg. Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie. Bonn 1996.
- Daniel, H.-D. / Hornbostel, S. „Das SPIEGEL-Ranking. Mediensensation oder ein Beitrag zur hochschulvergleichenden Lehrevaluation?“ Universität und Lehre. Hrsg. P. Mohler. Münster 1994: 29-44.
- Güdler, J., Sack, D. „Berufsfeld Wissenschaft. Zum Einfluss institutioneller Reputation auf die Platzierung von Nachwuchs-Soziologen“. Soziologielehre in Deutschland. Hrsg. H.M. Artus und M. Herfurth. Opladen 1996: 143-178.
- Hornbostel, S. / Daniel, H.-D. „Studienbedingungen in der Soziologie“. Soziologielehre in Deutschland. Hrsg. H.M. Artus und M. Herfurth. Opladen 1996: 11-58.
- Hornbostel, S. Wissenschaftsindikatoren. Bewertungen in der Wissenschaft. Opladen 1997.
- Hornbostel, S. „Der Uni-Test Europa des SPIEGEL: Infotainment oder Entscheidungshilfe?“ Uni-Test Europa. Wo sich das Studieren im Ausland lohnt. Hrsg. M. Doerry und J. Mohr. Hamburg 1998: 149 - 162.
- Hornbostel, S. „Das SPIEGEL-Ranking deutscher Hochschulen und die Folgen: Interaktionsprozesse zwischen Öffentlichkeit und Wissenschaft“. Die Eigenwilligkeit sozialer Prozesse Hrsg. J. Gerhards und R. Hitzler. Opladen 1999: 174-205.
- Hornbostel, S. „Evaluation und Ranking - Führen sie zu mehr Transparenz und Vergleichbarkeit?“ Beiträge zur Hochschulpolitik. Hrsg: Hochschulrektorenkonferenz. 4 1999: 81-96.
- Hornbostel, S. „Welche Indikatoren zu welchem Zweck: Input, Throughput, Output.“ Qualitätsförderung durch Evaluation? Ziele, Aufgaben und Verfahren im Wandel Hrsg. M. Röbbcke und D. Simon. Wissenschaftszentrum Berlin 1999: 99-103.
- Hornbostel, S. „Der Studienführer des CHE – ein multidimensionales Ranking“. Hochschulranking. Zur Qualitätsbewertung von Studium und Lehre. Hrsg. U. Engel. Frankfurt/M 2001a: 83-120.
- Hornbostel, S. „Third Party Funding of German Universities. An Indicator of Research Activity?“ Scientometrics Vol. 50 2001b: 523-537.
- Koch, J. „Wenn „mehr“ nicht gleichbedeutend mit „besser“ ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen“. ZUMA Nachrichten 42 1998: 66-90.
- Krempkow, R. Ist „gute Lehre“ meßbar? Untersuchungen zur Validität, Zuverlässigkeit und Vergleichbarkeit studentischer Lehrbewertungen. Marburg 1999.
- Kromrey, H. „Qualität und Evaluation im system Hochschule“. Evaluationsforschung (Hrsg. R. Stockmann). Opladen 2000.
- Lewin, K., Heublein, U., Schreiber, J., Sommer, D. „Vorbereitung auf das Studium und Informationsstand deutscher Studienanfänger bei Studienbeginn“. HIS Kurzinformation A8/1997.
- Meinefeld, W. „Hochschulranking. Eine unsichere Basis für Entscheidungen“. Forschung & Lehre 1 2000: 23-25.

Müller-Böling, D. / Hornbostel, S. „Fehlinterpretationen und Vorurteile. Vom Umgang mit Hochschulrankings und deren Nutzen“. *Forschung & Lehre* 2 2000: 81-38.

Ott, R. „Darstellung und Beurteilung von Hochschul-Rankings in Deutschland“. *Beiträge zur Hochschulforschung* 4 1999: 309-323.

Scheuch, E. K. „Wie gut sind unsere Hochschulen?“ *Soziologie* 2 1990: 73-90.

Schimank, U. *Hochschulforschung im Schatten der Lehre*. Frankfurt/M. 1995.

Süllwold, F. „Ranking ist oft ein Synonym für Unsinn“. *Forschung & Lehre* 11 1997: 578-579.

Weingart, P. „Forschungsindikatoren: Instrumente politischer Legitimierung oder organisatorischen Lernens“. *Qualitätssicherung in Hochschulen*. Hrsg. Detlef Müller-Böling. Gütersloh 1995: 73-84.

Zur Validität und zum Informationsgehalt von Studierendenbefragungen als Evaluation

Eine Reanalyse der CHE-Daten

Helmut Kromrey

Vorbemerkung

Der Zufall fügt es, dass Mitte Oktober 2000 in der Zeitschrift „Die Betriebswirtschaft“ ein Aufsatz erscheint, der sich grundlegend und umfassend mit Ansätzen zum Hochschul-Ranking - u. a. auch mit deren methodologischen Problemen - auseinandersetzt (Bayer, Christian R., 2000, s. auch den Beitrag des Autors in diesem Tagungsband, S. *** ff). Das eröffnet mir die Möglichkeit, mein Thema etwas genereller zu interpretieren als dies. zunächst vorgesehen war.

1. Zur Schwierigkeit angewandter Sozialforschung

Studierendenbefragungen zur Evaluation von Lehre und Studium (bzw. von Studienbedingungen) werden nicht zum Zwecke der Grundlagenforschung oder der Theorieentwicklung unternommen, sondern gehören zu den von Politik und Öffentlichkeit geforderten Beiträgen der Sozialwissenschaft zur gesellschaftlichen Praxis durch empirisch abgesicherte Informationen. Für Sozialwissenschaftler ist dies im Vergleich zur theoretisch orientierten Grundlagenforschung durchaus nicht die leichtere Aufgabe.

Jeder anwendungsorientierte Forscher kennt die Ansprüche, denen sich Sozialwissenschaftler in der Praxis gegenübersehen:

- Die Forschung soll: schnell vonstatten gehen, also hochaktuelle Ergebnisse liefern; sie soll dabei wenig Aufwand erfordern; manchmal soll sie sogar nach erster fachmännischer Anleitung vom Laien selbst weiterführbar sein.
- Die Ergebnisse sollen unmittelbar anwendungsrelevant sein, sie sollen die soziale Realität überschaubar und durchschaubar darstellen; d. h. sie sollen die Welt auf die (vermeintlich) wesentlichen Dinge reduzieren und dies möglichst allgemein verständlich.
- Dennoch sollen die Resultate selbstverständlich valide bleiben, also nicht ein durch zu große Vereinfachung verzerrtes Bild zeichnen. Im Unterschied zur Grundlagenforschung darf sich anwendungsorientierte Forschung, da ihre Ergebnisse unmittelbar handlungsrelevant werden sollen, nicht irren.

Die Versuche des Rankings von ganzen Hochschulen (früher) bzw. von Fachbereichen oder Studiengängen (auch von Lehrenden in den Hochschulen oder Fachbereichen) sind ein Beispiel für solche Anforderungen.

Diesen Ansprüchen steht hinderlich entgegen, dass soziale Sachverhalte immer komplex sind, häufig nur dem Fachmann verständlich, schwer durchschaubar, selten direkt beeinflussbar.

Was kann bei dieser Konstellation passieren? Es können - vielleicht sogar: es müssen - sich in mehrfacher Hinsicht Fehler einschleichen:

- Fehler bei der Informationssammlung durch Orientierung an der Forderung nach Zeit- und Aufwandminimierung (schnell, einfach, laienverträglich);
- Fehler bei der Informationsaufbereitung (möglichst ‚einfache Statistik‘, Reduktion der Informationsfülle auf wenige ‚Kennwerte‘, Verzicht auf komplexe Modelle);
- Fehler bei der Interpretation der Daten sowie bei Schlussfolgerungen und Empfehlungen (empirische Daten werden uninterpretiert für ‚Fakten‘ genommen, ‚mehr‘ wird mit ‚besser‘ gleichgesetzt; man orientiert sich an leichter Durchsetzbarkeit, an einfachen Handlungsempfehlungen, an der Sichtbarkeit des Eingreifens - Motto: Es muss etwas geschehen!).

Anwendungsorientierte Sozialforschung ist also keine einfache Sache. Das wird jedoch nicht überall so gesehen, was ein weiteres Problem mit sich bringt: Wenn direkt anwendungsorientierte Informationssammlung als die anspruchslosere Variante gilt, dann - so eine verbreitete Vorstellung - kann das auch jeder methodisch Halbgebildete, dann kann das jeder, der irgendwann einmal etwas von empirischen Methoden gehört hat; dann können das auch diejenigen, die den Methodenpflichtteil ihrer Ausbildung mit der Einstellung über sich haben ergehen lassen:

- so wenig wie möglich davon, nur das ‚wirklich Notwendige‘, insbesondere:
- einfache Verfahren der Informationserhebung,
- einfache Verfahren der Auswertung, nur ganz ‚einfache Statistik‘ (wenn überhaupt),
- nur das, was absolut unumgänglich ist; alles andere ist allenfalls etwas für ‚Methoden-Freaks‘.

Dies steht nicht nur im Widerspruch zu der gerade skizzierten Gefahr, an vielen Stellen viele Fehler zu begehen. Es beschwört auch eine ganz besondere Gefahr herauf - nämlich dass die Gefahren methodischer Halbbildung als solche überhaupt nicht erkannt werden: Je schlichter die methodische Ausbildung, um so überzeugter ist man von der Richtigkeit (wenn nicht gar: ‚Objektivität‘) seiner Resultate. Um so weniger besteht dann Anlass für Zweifel an der schönen Gleichung ‚einfach = gut‘.

2. Evaluation als Spezialfall angewandter Sozialforschung

Ich habe zu Beginn drei Bereiche genannt, in denen typischerweise gravierende Fehler unterlaufen können (bzw. müssen):

- bei der Informationssammlung,
- bei der Informationsaufbereitung bzw. -auswertung und

- bei den Schlussfolgerungen und Empfehlungen.

Die Konsequenzen solcher Fehler sind besonders eklatant bei jenem Typ von angewandter Sozialwissenschaft, der hier zur Debatte steht: der Evaluation (in allen ihren Schattierungen von Evaluationsforschung bis zur Politikberatung). Seit den bekannten SPIEGEL-Rankings (Welche Uni ist die beste?) hat die Empirie ihren festen Platz auch in der Diskussion um die Hochschulen und in den Hochschulen. Spätestens seit diesem Zeitpunkt sind auch die Bundes- und Landesministerien wieder auf die Sozialwissenschaftler als anwendungsorientierte Forscher aufmerksam geworden und erwarten von ihnen nützliche Beiträge:

- So heißt es zum Beispiel im Aktionsprogramm Qualität der Lehre der nordrhein-westfälischen Landesregierung 1991: „Im Mittelpunkt einer Kritik des Ausbildungsangebots steht die Beurteilung der Lehrveranstaltungen und damit auch der Lehrenden (Veranstaltungskritik). Daneben kommt die Beurteilung des Prüfungsbetriebs und - mit gewissen Einschränkungen - der Studiengänge in Betracht“ (MWF NW, 1991: S. 124).
- Und 1992 forderte der seinerzeitige Bundesbildungsminister Ortleb ein ‘Anknüpfen von "Belohnungsmechanismen" an gute Leistungen in der Lehre oder die systematische Evaluation von Lehrleistungen der Hochschulen’ - so eine seiner ‘Thesen zur Belebung der Leistungskraft der Hochschulen’ (Informationen Bildung Wissenschaft, Nr. 7-8/92: S. 90–92).
- Ein weiterer Diskussionsstrang läuft an den Hochschulen selbst darauf hinaus, die knapp gewordenen (und auch künftig knapper werdenden) Finanzmittel „leistungsbezogen“ zuzuteilen.
- Schließlich sollen zwischen den (öffentlich-rechtlichen) Hochschulen Qualitätskennziffern so etwas wie einen (privatwirtschaftsanalogen) Wettbewerb simulieren, soll das Handeln dieser Einrichtungen an Kunden- oder Klientenorientierung gewinnen: der erreichte Platz auf der Ranking-Liste als Ersatz für die Börsennotierung!
- Da eine zentrale Kunden- oder Klientengruppe die Studierenden sind, ist eine zwangsläufige Konsequenz, dass diese für ein marktgerechtes Nachfrageverhalten die notwendige Transparenz benötigen. Diese soll durch Informationen über die Angebote und Anbieter der Dienstleistung Lehre hergestellt werden: der CHE-Studienführer als die Financial Times der Studierenden!

Gefragt sind in allen diesen Fällen empirische und damit - so hofft man - quasi objektive Bewertungen nicht nur der Lehre, sondern des gesamten Ausbildungs- und Prüfungsbetriebs.

3. Studierendenbefragung als Evaluationsinstrument

Selbstverständlich verfügt die anwendungsorientierte Sozialforschung über ein geeignetes Design zur empirischen Bewertung (auch von Lehre und Ausbildung), nämlich das Design der

Programm-Evaluation. Es ist allerdings methodisch aufwendig, verlangt vom Durchführenden Forschungserfahrung, kostet viel Zeit, setzt u. a. relativ genaue Kenntnisse der Ziele, Instrumente und Zielgruppen des Programms sowie den Konsens über geeignete Beurteilungskriterien voraus und muss selbstverständlich auch auf die Inhalte des zu bewertenden Programms zugeschnitten sein. Es ist also alles andere als ein einfaches Verfahren.

An ein solches differenziertes Forschungsprogramm ist denn auch gar nicht gedacht, wenn im hochschulpolitischen Kontext der Ruf nach Evaluation der Qualität von Lehre und Ausbildung laut wird. Gedacht wird nicht an gegenstandsspezifische, sondern an möglichst generelle Kriterien für ‚Qualität‘, an Kriterien, die gerade stoff- und fach- und lernzielunabhängig sind, die - bei universitätsinterner Perspektive - möglichst flächendeckend zum gleichen Zeitpunkt in der gesamten Universität anwendbar sein sollen. Beim Hochschulranking geht der Anspruch noch etwas weiter: Die Kriterien sollen über alle Hochschulen hinweg bundesweit anwendbar sein.

Sagt dazu der verantwortungsbewusste und methodisch reflektiert verfahrenende Sozialforscher, diese Aufgabe sei - wenn wirklich Qualität gemessen und danach ranggeordnet werden soll - unlösbar, so scheint demgegenüber für den methodisch Halbgebildeten die Lösung ganz einfach: Man befrage - so der auf den ersten Blick durchaus plausibel erscheinende Rat - die Studierenden in den jeweiligen Organisationseinheiten und erhebe deren Bewertungen; und man tue dies anhand kurzer Fragebögen mit relativ globalen (also: inhaltsunabhängigen) Beurteilungen. Die Studierenden - dieses Argument liegt auf der Hand - sind als die vom Lehr- und Ausbildungsangebot ganz konkret ‚Betroffenen‘ am besten in der Lage, aus eigener Erfahrung die Qualität des ihnen Gebotenen zuverlässig zu beurteilen.

Zwar ist ein solches Vorgehen - methodisch gesehen - kein Evaluationsdesign, sondern ‚ganz normale‘ Umfrageforschung. Dies wäre jedoch kein Problem, wenn die folgenden, in den Materialien eines hochschuldidaktischen Zentrums zu lesenden Ausführungen zuträfen:

- Urteile (Schätzungen) von Studenten über die Lehre sind verlässlicher als Urteile der Dozenten über die Leistung der Studenten und ebenso verlässlich wie Urteile von Kollegen über die Lehre,
- sie legen praktisch dieselben Kriterien für gute Lehre an wie die Dozenten selbst,
- sie sind - wenn man etwa 20-30 Studenten urteilen lässt - zuverlässig wie professionelle Testverfahren,
- sie sind von anderen Merkmalen der Studenten selbst und der Dozenten wenig beeinflusst (Schmidt, Jörn, 1980: S. 51–52).

Obwohl bereits runde 20 Jahre alt, wurden und werden diese Argumente immer wieder vertreten. Leider jedoch ist jede einzelne dieser Aussagen falsch (allenfalls mit Ausnahme der zweiten). Doch ist dies nicht ohne weiteres erkennbar - insbesondere dann nicht, wenn man auf die gewünschten ‚einfachen Verfahren‘ der Datenerhebung und -auswertung zurückgreift. Zwar lernt man als Studierende/r schon in den ersten Semestern manches von den Gefahren und Fallstricken der empirischen Sozialforschung (insbesondere des Instruments Befragung), vergisst es dann aber erfolgreich wieder. Es ist erstaunlich, wie viel von ihrem theoretischen

Wissen selbst gestandene Soziologen vergessen können, sobald es um die praktische Anwendung geht.

So gibt es denn auch ganz aktuelle Beispiele dafür, dass Sozialwissenschaftler Befragungsinstrumente vorlegen, die das im obigen Zitat erhoffte Wunder vollbringen sollen. Etwa ein von Psychologen entwickeltes und 1994 veröffentlichtes ‚Heidelberger Inventar zur Lehrveranstaltungs-Evaluation‘ (abgekürzt: HILVE) oder ein im ‚Projekt Pro Lehre‘ der Freien Universität Berlin konzipiertes und wiederholt eingesetztes ‚Studienbarometer‘, das zu vergleichenden Aussagen über die Ausbildungsqualität an Fachbereichen und Instituten beitragen soll und dabei mit einer einzigen Seite auskommt.

Diese allgemeinen kritischen Vorbemerkungen beziehen sich - ich möchte das ausdrücklich hervorheben - nicht auf den vom CHE vorgelegten Studienführer und die darin vorgenommenen Rangordnungen. Die Bearbeiterinnen und Bearbeiter des Studienführers haben die Kritik an den Ranking-Vorläufern von SPIEGEL bis FOCUS zur Kenntnis genommen und vermeiden viele der dort begangenen Fehler. Sie setzen insbesondere nicht mehr die Einschätzungen von Studierenden mit „Evaluation“ im Sinne einer quasi-objektiven Qualitätsmessung gleich, sondern präsentieren je Studiengang eine ganze Palette objektiver und subjektiver Daten, unter denen die Studierendenbefragung nur zwei von neun Bausteinen liefert (Buhr, Petra; Giebisch, Petra; Hombostel, Stefan; Müller-Böling, Detlef, 2000: S. 2ff., 10ff.).

Da ich aber zu einer Reanalyse der Studierendenbewertungen aus der CHE-Erhebung eingeladen wurde, habe ich natürlich nach Belegen in dieser Studie fahnden müssen.

4. Fehler bei der Informationssammlung

Ich komme damit zum angekündigten Beispiel für die erste Fehlerquelle - **‘Fehler bei der Informationssammlung’**. Dazu zunächst ein Exkurs zu der schon kurz angesprochenen Evaluations-Befragung, die regelmäßig an der FU Berlin durchgeführt und ‚Studienbarometer‘ genannt wird. Eines der Ziele des ‚Studienbarometers‘ ist es, „... schnell und systematisch die Einschätzung von Studierenden zu wichtigen Aspekten des Studienbetriebs am Fachbereich“ zu erheben. „Diese sollten ... dann zum Ausgangspunkt von konkreten Maßnahmen zur Verbesserung der Verhältnisse“ gemacht werden (PPL-Schreiben vom 5. Oktober 1995). U. a. wird auch ein hochschulinternes Ranking von Fachbereichen und Instituten angestrebt. Zu bewerten sind von den befragten Studierenden für Ihr Institut bzw. Ihren Fachbereich so allgemeine Statements wie:

- die didaktische Kompetenz der Lehrenden
- die Vollständigkeit des Lehrangebots
- die Übersichtlichkeit des Lehrangebots
- etc.

In Seminaren zum Thema „Befragung“ nehme ich den ‚Studienbarometer‘-Fragebogen immer gern als Material für eine kleine Übung: Die Teilnehmer füllen zunächst entsprechend der dort

gegebenen Anleitung den Bogen aus; anschließend sollen sie die von ihnen gelesenen Fragen sowie ihre in der Beurteilungsskala eingetragenen Kreuzchen in eigenen Worten verbalisieren. Es ist für die Teilnehmer immer wieder frappierend, wie viele unterschiedliche Fragen ein einziges, scheinbar einfaches und klares Item repräsentieren kann. Eines der Items fragt nach dem „Praxisbezug der Lehre“. Die genannte Übung illustriert: „Praxisbezug der Lehre“ kann für die Studierenden der Soziologie sein: die Frage nach dem Bezug zu dem von ihnen später individuell angestrebten Beruf (Problem: Ich weiß noch nicht, was ich später tun will) oder auch - Soziologen müssen immer abstrahieren - die Vermittlung von Kenntnissen und Fertigkeiten über sozialwissenschaftliche Tätigkeitsfelder generell (Problem: Welche Berufstätigkeiten für Soziologen gibt es eigentlich?). „Praxisbezug“ kann aber auch etwas ganz anderes sein: nämlich die Illustration des Lernstoffs durch Beispiele, entweder Beispiele „aus der Praxis“ oder fiktive Beispiele zum Einüben des zu Lernenden (das Berechnen von Mittelwerten oder Korrelationen in der Statistik-Übung, die Entwicklung eines Fragebogens in der Methodenlehre-Übung ist dann Praxisbezug). Wie steht es nun bei solcher Unterschiedlichkeit von Fragen, die dieses eine Item repräsentieren kann, mit der Einlösung des angestrebten Zwecks der Befragung, „zum Ausgangspunkt von konkreten Maßnahmen zur Verbesserung der Verhältnisse“ zu werden? – Es kann nicht eingelöst werden, da durch kein Analyseverfahren rekonstruierbar ist, worauf das konkrete Kreuzchen im konkreten Fragebogen eine Antwort ist.

Der CHE-Fragebogen verwendet ganz ähnliche Statements zur Beurteilung des Lehrangebots im Studiengang (Frage 4 im Studierenden-Fragebogen; s. Buhr u.a., 2000: Anhang 3.) anhand einer Notenskala von 1 (sehr gut) bis 6 (sehr schlecht), u. a. auch die Frage nach dem „Praxisbezug der Lehrveranstaltungen“. Natürlich weiß ich nicht, was die Studierenden in den Fachhochschulen, Universitäten und Technischen Hochschulen in den verschiedensten Studiengängen damit assoziiert haben; aber es scheint zumindest auch auf die Doppelbedeutung „Relevanz für die außeruniversitäre Praxis“ und „gegenständliche, illustrative Lehre“ hinauszuweisen. Eine Dimensionsanalyse der Antworten (statistisches Modell: Faktorenanalyse) zeigt, dass „Praxisbezug der Lehrveranstaltungen“ und „Angebot von Lehrveranstaltungen durch Praktiker“ auf derselben Dimension liegen wie die „didaktische Vermittlung des Lehrstoffs“ sowie die „Abstimmung des Lehrangebots auf Prüfungsanforderungen“. Eine zweite Dimension repräsentiert den Forschungsbezug, kombiniert mit inhaltlicher Breite des Lehrangebots, Interdisziplinarität und Internationalität.¹

Auf die üblichen Kritikpunkte an der Vorgabe derart globaler Bewertungs-Items angesichts eines heterogenen Beurteilungs-Gegenstands (es gibt nicht „die“ Lehre und nicht „die“ Didaktik) und angesichts unterschiedlicher Grade des selektiven Informiertseins der Beurteiler will ich hier nicht eingehen. Sie gelten generell für jede Evaluierung durch Befragung. Sie ließen sich zwar durch geeignete Zusatzfragen zum „Expertenstatus“ der Befragten und zu ihrer Interpretation der Fragen ausräumen; dies würde jedoch gerade die gewünschte Einfachheit der Erhebung konterkarieren.

¹ Nebenbei bemerkt: Auch das Item „Angebot an Projektseminaren und Praktika in der Lehre“ scheint misslungen, da es zwei konkurrierende Bewertungsdimensionen (Projektseminare=Forschung und Praktika=Praxisbezug) in einen Topf wirft. Konsequenterweise weist es mittelhohe Ladungen auf beiden Faktoren auf.

5. Fehler bei der Informationsaufbereitung

Als zweite Fehlerquelle hatte ich zu Beginn ‚Fehler bei der Informationsaufbereitung‘ durch zu einfache Aufbereitung genannt. Nicht ausmerzbar scheint insbesondere die Vorliebe für die Berechnung des arithmetischen Mittels der erhobenen Studenturteile zu sein. In Kauf genommen wird hier die Gefahr, Daten über ein Erhebungs-konstrukt (= ein Befragungs-Artefakt), genannt „die Lehre“ zu sammeln und zu analysieren. Hier kann ich nun ohne Umweg direkt zum CHE-Studienführer kommen.

Wie in anderen Erhebungen, so zeigt sich auch hier: Die befragten Studierenden sind sich natürlich nicht einig in ihren Einschätzungen. Derselbe Sachverhalt (der identische Situationsaspekt in derselben Hochschule) wird von einem Teil als ganz schlecht, von einem anderen als mittelmäßig, von noch einem anderen Teil als gut bis sehr gut wahrgenommen. Dieser Befund aber ließe sich weder in einfacher Weise vermitteln; noch ließe sich so der gewünschte numerische Vergleich zwischen den jeweiligen Studienbereichen an Universitäten und Fachhochschulen vornehmen. Also wird auf einfache Statistik zurückgegriffen: Ein Durchschnittswert muss her (im allgemeinen eben das aus dem Alltag vertraute arithmetische Mittel²): Wird bei unzulässiger Vereinfachung durch die Tendenz zur Pauschalisierung in den gestellten Fragen die Gefahr der Produktion von Erhebungsartefakten in Kauf genommen (z. B. ‚didaktische Kompetenz der Lehrenden am Fachbereich‘), so tritt nun zusätzlich ein statistisches Artefakt³ hinzu: ‚Urteil der Studierenden am Fachbereich‘. Dieses ermöglicht, Urteilsdurchschnitte pro Studienbereich darzustellen und Rangordnungen der Hochschulen anhand der jeweiligen Einschätzungsaspekte zu bilden.

Die Bearbeiterinnen und Bearbeiter des CHE-Studienführers kennen zwar diese Fundamentalkritik. Sie berechnen dennoch weiterhin solche Mittelwerte und ordnen pro Studiengang die einzelnen Hochschulen nach der zweiten Kommastelle. Aber sie wollen das - eben wegen der Problematik dieses Vorgehens - nicht so ernst genommen wissen. Übrig bleibt bei ihnen nach statistischer Berücksichtigung der Streuung der Antworten lediglich die grobe Einordnung der Hochschulen in eine Spitzengruppe, einen unspezifischen Mittelbereich und eine Schlussgruppe. Als studentische Qualitätsaussage interpretiert wird von ihnen lediglich die Zugehörigkeit eines Studienbereichs zur Spitzen- oder zur Schlussgruppe. Ein Großteil der mit viel Erhebungsaufwand gewonnenen empirischen Informationen geht so verloren; oder anders formuliert: wird wegen der Wahl eines methodisch fragwürdigen statistischen Modells ungenutzt „vertan“. (Darauf komme ich später noch einmal zurück.)

² Dabei wird bei allen mir bekannten Darstellungen von Ergebnissen solcher ‚Lehrevaluationen‘ die Eignung des arithmetischen Mittels als Instrument der Informationsreduktion nicht einmal ansatzweise diskutiert, sondern - wie im obigen Zitat aus dem Text des Hochschuldidaktischen Zentrums - als im gegebenen Zusammenhang selbstverständlich erfüllt unterstellt.

³ Ein statistisches Artefakt ist das arithmetische Mittel dann, wenn wesentliche Voraussetzungen für die Anwendung des Modells nicht erfüllt sind. Hier trifft das immerhin für zwei Voraussetzungen zu. Zum einen weisen die Daten lediglich Ordinalskalenniveau (statt des messtheoretisch geforderten metrischen Niveaus) auf: Das ‚sehr gut‘ (linker Skalen-Endpunkt) wird in die Ziffer ‚1‘ codiert, das ‚sehr schlecht‘ (rechter Skalen-Endpunkt) in die Ziffer ‚6‘; für Zwischenwerte stehen die Ziffern 2, 3, 4 und 5. Eine zweite Voraussetzung liegt in einer Verteilungsannahme des Modells arithmetisches Mittel begründet: Die Messwerte-Verteilung muss, damit der Mittelwert empirisch sinnvoll interpretiert werden kann, eine zentrale Tendenz aufweisen. Auch diese Voraussetzung ist bei Zufriedenheitsbefragungen häufig empirisch nicht erfüllt.

Zwar ließe sich auch zur Qualität der erhobenen Daten selbst einiges anmerken; insbesondere täuscht die Anwendung wahrscheinlichkeitsstatistischer Signifikanzmodelle eine Objektivität der Gruppierung vor, die wegen fehlender Zufallsstichprobe⁴ absolut nicht gegeben ist. Das möchte ich aber hier nicht zum Thema machen; statt dessen möchte ich auf Probleme der von den Bearbeitern bevorzugten abstrahierenden Orientierung an Variablen und daraus gebildeten Indizes hinweisen.

Auch die Bildung von Indizes impliziert die oben schon aus einer hochschuldidaktischen Schrift zitierte Erwartung, individuelle Wahrnehmungen und Urteile würden sich bei hinreichend großer Zahl von Informationen „ausmitteln“. Dieses Argument setzt voraus, dass die Befragten sich in ihren Urteilen an einer einzigen latenten Dimension orientieren (die man sich vorstellen kann als „Qualität“, und zwar verstanden als Eigenschaft des zu beurteilenden Sachverhalts). Wäre diese Annahme zutreffend, gäbe es selbstverständlich keine systematischen Abhängigkeiten von Merkmalen des Urteilers und seines Bezugs zum Gegenstand seiner Bewertung; in diesem Fall wäre im übrigen auch die Mittelwertberechnung pro Variable ein geeignetes statistisches Modell. Dem ist aber nicht so. Alle mir bekannten differenzierteren Auswertungen zeigen, dass jedes subjektive Qualitätsurteil eine Aussage über die Relation Urteilender/Gegenstand ist - und diese Relation kann natürlich systematisch zwischen den Urteilenden variieren. D. h. es kann bei (objektiv) identischem Bewertungsobjekt Teilgruppen von Befragten geben, die (subjektiv) unterschiedliche Sachverhalte beurteilen; und das kann man nicht „ausmitteln“. Berechnet man dennoch aus einer größeren Zahl von Items einen zusammenfassenden Index über alle Befragten hinweg, erhält man zwar eine eindimensionale Messgröße, aber leider entspricht dieser kein empirisch gemessener Sachverhalt - auch sie ist in diesem Fall ein Auswertungsartefakt.

Sofern man nun nicht über Informationen zur Relation Urteilender/Gegenstand verfügt - und das ist bei standardisierten Befragungen meist der Fall, weil sie nicht erhoben werden -, kann man immerhin versuchen, sie hypothetisch zu rekonstruieren, indem man nach empirisch voneinander unterscheidbaren Bewertungsmustern der Befragten sucht (um z. B. die „Fans“ zu identifizieren, die alles ganz toll finden; oder auch ihr Gegenstück, die „Fundamentalkritiker“, die an allem kein gutes Haar lassen. Dazwischen liegen Gruppen, die Teilbereiche gut, andere Teilbereiche weniger gut oder schlecht finden. Geschieht dies bei jeweils identischem Bewertungsobjekt, kann auf die Situation der Urteilenden geschlossen werden (wie man dies ja auch im Alltag versucht, wenn man von verschiedenen Personen stark voneinander abweichende Aussagen zum selben Sachverhalt erhält). Ein geeignetes statistisches Modell für diese Rekonstruktion ist die Analyse latenter Strukturen (LCA); eine gute deskriptive Annäherung ist die - einfacher zu handhabende - Clusteranalyse. An die Stelle der Suche nach Ähnlichkeiten zwischen Variablen-Verteilungen (wie bei Korrelations- und Regressionsmodellen oder bei der Faktorenanalyse) tritt hier die Suche nach Ähnlichkeiten zwischen Objekten (hier: Befragten und ihren Urteilsprofilen).

Bei variablenorientierter Vorgehensweise ist es guter Brauch, von einer Dreiteilung der Welt auszugehen: durchschnittlich - besser als der Durchschnitt - schlechter als der Durchschnitt

⁴ Bei nur gut 20% Rücklauf degeneriert jede geplante Wahrscheinlichkeitsauswahl zu einer realisierten Selbstselektion.

(dieses Prinzip finden wir auch in der CHE-Gruppierung der Hochschulen). Gibt man analog für die Suche nach Ähnlichkeiten zwischen Personen auch eine Dreier-Lösung vor (bei der Clusteranalyse also eine Drei-Cluster-Lösung), dann lässt sich in den Daten in der Tat mit hoher statistischer „Güte“⁵ eine solche Gruppierung von Befragten vollziehen: solche, die alles ziemlich gut beurteilen - solche, die bei jedem Item eine kritischere Sicht auf ihre Hochschule zum Ausdruck bringen - und der große mittlere Bereich mit der tendenziellen Aussage: „es geht so“.

Studentische Urteile über ihren Studiengang (Fragen 4 und 14; 1=sehr gut, 6=sehr schlecht):	Cluster				F-Statistik
	1 alles gut	2 mittel	3 alles schlecht	Gesamt	
Lehrangebot inhaltliche Breite	1,8	2,4	3,2	2,3	4055
Lehrangebot Projektseminare etc.	2,1	3,0	4,1	2,9	6453
Studienorg. Abstimmung Prüfung	2,0	2,5	3,4	2,5	4212
Lehrangebot Forschung	2,6	3,8	4,6	3,5	5419
Lehrangebot Praxisbezug	2,0	2,9	4,2	2,8	9156
Lehrangebot Zugang Pflichtveranst.	1,3	1,6	2,2	1,6	1365
Lehrangebot internationale Ausrichtung	2,4	3,5	4,3	3,2	5333
Lehrangebot Interdisziplinarität	2,2	3,1	4,0	2,9	5645
Lehrangebot Vermittlung Lehrstoff	2,3	3,1	4,1	3,0	7177
Lehrangebot Praktiker-Veranstaltungen	2,1	3,0	4,3	2,9	6976
Gesamturteil Studiensituation	1,8	2,5	3,6	2,4	10686
	N ¹	8385	10125	4503	23013

Tabelle 1: Clusteranalyse Studierenden-Urteile zum Studium (CHE-Daten; drei Urteilsprofile)

Unterstellt wird bei der Erhebung von Evaluationen per Befragung - wie im Abschnitt III bereits dargestellt -, dass die abgegebenen Urteile Eigenschaften des in der Frage benannten Sachverhalts widerspiegeln und nicht systematisch mit Merkmalen der urteilenden Befragten zusammenhängen (s. das dort wiedergegebene Zitat aus einer hochschuldidaktischen Veröffentlichung). Doch schon bei der hier präsentierten groben Dreier-Gruppierung zeigen Kreuztabellierungen, dass dem nicht so ist.

⁵ Die berechneten statistischen F-Werte belegen für alle Variablen eine außerordentlich hohe Trennschärfe zwischen den Gruppen: Die Zwischengruppenvarianz ist (mit Ausnahme der Beurteilung der Zugangsmöglichkeiten zu den Pflichtveranstaltungen, die generell als gut bis sehr gut eingeschätzt wird) durchweg mehrtausendfach höher als die Varianz innerhalb der drei Gruppen.

"Bias-Variablen":		Cluster			Gesamt
		1	2	3	
Geschlecht	weibl.	18,2%	24,6%	27,7%	22,9%
	männl.	81,8%	75,4%	72,3%	77,1%
Alter	bis 24 J.	49,6%	42,4%	29,9%	42,6%
	bis 29 J.	43,6%	48,7%	52,2%	47,5%
	30 u.m.	6,8%	8,9%	17,9%	9,9%
Schulnote	sehr gut	7,9%	6,2%	5,7%	6,7%
	gut	53,9%	50,1%	48,7%	51,2%
	befr.	36,6%	41,4%	43,1%	40,0%
	ausr.	1,6%	2,3%	2,5%	2,1%
Fachsemester	bis 8 Sem.	71,0%	65,4%	49,6%	64,4%
	bis 11 Sem.	22,4%	25,3%	31,2%	25,4%
	bis 15 Sem.	5,1%	7,1%	13,5%	7,6%
	16 u.m.	1,5%	2,1%	5,7%	2,6%
Vollzeitstudium	nein	14,6%	21,8%	36,1%	22,0%
	ja	85,4%	78,2%	63,9%	78,0%
Job außerh. der HS	nein	36,1%	31,1%	25,2%	31,8%
	ja	63,9%	68,9%	74,8%	68,2%
Job an der HS	nein	74,8%	80,5%	82,5%	78,8%
	ja	25,2%	19,5%	17,5%	21,2%

Tabelle 2: Zusammenhang subjektiver Evaluationen mit Merkmalen der Evaluierenden

Die abgegebene Evaluation wird nicht allein vom Beurteilungsgegenstand bestimmt, sondern korreliert systematisch mit Merkmalen der bewertenden Studierenden: So urteilen Frauen kritischer als Männer, ältere Befragte kritischer als jüngere, Personen mit schlechten Schulnoten und solche in höheren Semestern nehmen die Studienbedingungen negativer wahr als diejenigen mit guten Schulnoten und in niedrigeren Semestern, Vollzeitstudierende ebenso wie Personen mit einem Job an der Hochschule urteilen positiver als Befragte mit Jobs außerhalb der Hochschule und Teilzeitstudierende.⁶ Zwar ist jede der Korrelationen für sich genommen nicht besonders hoch; aber im Datensatz sind lediglich die eher urteilsneutralen Merkmale der Personen enthalten. Informationen über die als Urteilsmaßstab relevanteren Merkmale wie Interesse / intrinsische Motivation, Anspruchsniveau, Arbeits-/Lernstil, Berufsziel etc. wurden leider nicht erhoben.

⁶ Häufig wird die Existenz des Einflusses solcher sog. Bias-Variablen aufgrund statistischer Analysen verneint, die eine lineare Abhängigkeit unterstellen. Ein Blick auf die Tabelle 2 zeigt, dass ein solcher linearer

Kommt eine Dreiergruppierung wie die eben vorgestellte zwar dem Wunsch nach Einfachheit und Übersichtlichkeit entgegen, so verbindet sich damit doch immer auch die Gefahr der Übereinfachung: Von einer dominierenden Dimension (hier: positiv/negativ) können hoch relevante Differenzierungen verdeckt werden. Dies trifft in unserem Fall für das mittlere der drei Urteilsprofile zu, in dem zwei qualitativ extrem unterschiedliche Arten von Kritik an Teilaspekten des Studienangebots zusammenfallen. Schon bei einer Vorgabe von vier Clustern wird erkennbar, dass die in der Faktorenanalyse sich andeutenden beiden konkurrierenden Perspektiven Forschungsbezug versus Praxisbezug sich auf der Personenebene zu zwei voneinander abgegrenzten Urteilsprofilen konkretisieren (die lediglich „im Durchschnitt“ gleich sind: Gesamturteil Studiensituation = 2,4 bzw. 2,5).

Studentische Urteile über ihren Studiengang (Fragen 4 und 14; 1=sehr gut, 6=sehr schlecht):	Cluster				Gesamt	
	1 alles gut	2 Kritik Praxis	3 Kritik Forschung	4 alles schlecht		
Lehrangebot inhaltliche Breite	1,8	2,2	2,6	3,2	2,3	
Lehrangebot Projektseminare etc.	2,0	3,0	3,0	4,2	2,9	
Studienorg. Abstimmung Prüfung	1,9	2,6	2,4	3,5	2,5	
Lehrangebot Forschung	2,6	3,2	4,3	4,6	3,5	
Lehrangebot Praxisbezug	1,9	3,2	2,5	4,3	2,8	
Lehrangebot Zugang zu Pflichtveranst.	1,3	1,6	1,6	2,2	1,6	
Lehrangebot internat. Ausrichtung	2,3	2,8	4,1	4,4	3,2	
Lehrangebot Interdisziplinarität	2,2	2,8	3,3	4,1	2,9	
Lehrangebot Vermittlung Lehrstoff	2,3	3,2	3,0	4,2	3,0	
Lehrangebot Praktiker-Veranstaltungen	2,0	3,7	2,4	4,3	2,9	
Gesamturteil Studiensituation	1,8	2,4	2,5	3,7	2,4	
	N	7556	5911	5562	3984	23013

Tabelle 3: Clusteranalyse Studierenden-Urteile zum Studium (CHE-Daten; vier Urteilsprofile)

Zwischen den beiden stabilen Urteilmustern „alles gut“ und „alles schlecht“ kristallisiert sich zum einen eine Tendenz zu (gemäßiger) Kritik insbesondere am Praxisbezug und an der Art der Vermittlung des Lehrstoffs heraus⁷, zum anderen tritt eine (deutlich) negative Benotung von Forschungsbezug und internationaler Ausrichtung des Lehrangebots hervor (alles andere erscheint als gut bis befriedigend).

Zusammenhang in der Tat bei keinem der genannten Merkmale der Befragten zuträfe. Die Gleichsetzung eines nicht vorhandenen linearen Einflusses mit „kein Einfluss“ führt also in die Irre.

⁷ Zwar erhält auch der Forschungsbezug nur eine „befriedigende“ Benotung; diese fällt jedoch immer noch etwas besser als im Gesamtdurchschnitt aus.

Weitergehende Aufgliederungen der individuellen Urteile (anhand der von CHE eingesetzten Batterie von Items) erhöhen zwar den Differenzierungsgrad, lassen aber keine neuen latenten Dimensionen für die Clusterbildung relevant werden. Bei - beispielsweise - einer empirischen Klassifizierung von sechs Gruppen bleiben zwischen den extremen Profilen „alles gut“ und „alles schlecht“ auch die beiden Kritikansätze am Praxisbezug und am Forschungsbezug erhalten; sie tauchen allerdings jeweils einmal in einer gemäßigten (Cluster 2 und 3) und einmal in einer ausgeprägt negativen Version auf (Cluster 4 und 5):

Studentische Urteile über ihren Studiengang (Fragen 4 und 14; 1=sehr gut, 6=sehr schlecht):	Cluster						
	1 alles gut	2 gut / Kritik Praxis	3 gut / Kritik Forschg	4 mittel / Kritik Praxis	5 mittel / Kritik Forschg	6 alles schlecht	
Lehrangebot inhaltliche Breite	1,7	2,0	2,3	2,6	2,8	3,5	
Lehrangebot Projektseminare etc.	1,8	2,6	2,6	3,7	3,4	4,4	
Studienorg. Abstimmung Prüfung	1,8	2,3	2,2	3,1	2,5	3,8	
Lehrangebot Forschung	2,3	2,7	3,8	3,8	4,5	4,9	
Lehrangebot Praxisbezug	1,8	2,8	2,2	3,9	2,9	4,6	
Lehrangebot Zugang Pflichtveranst.	1,2	1,5	1,4	1,9	1,8	2,4	
Lehrangebot internat. Ausrichtung	2,1	2,7	3,2	3,1	4,7	4,8	
Lehrangebot Interdisziplinarität	2,0	2,6	2,8	3,2	3,7	4,4	
Lehrangebot Vermittlung Lehrstoff	2,1	2,9	2,6	3,6	3,2	4,6	
Lehrangebot Praktiker-Veranstaltg.	1,8	3,4	2,0	4,1	2,9	4,6	
Gesamturteil Studiensituation	1,6	2,1	2,2	2,9	2,9	4,1	
	N	4344	4599	5005	3491	3552	2022

Tabelle 4: Clusteranalyse Studierenden-Urteile zum Studium (CHE-Daten; sechs Urteilsprofile)

Mit der Unterscheidung von vier Urteilsprofilen ist also in diesem Datensatz eine Lösung gefunden, die einerseits eine übersichtliche und gut interpretierbare Vereinfachung anbietet, die dabei andererseits aber keine systematische Verzerrung der empirischen Informationen - kein Auswertungsartefakt - produziert.

Statistische Auswertungen sollten natürlich kein Selbstzweck sein. Daher stellt sich die Frage: Ist diese über eine Drei-Gruppen-Lösung hinausgehende Differenzierung der Urteilsprofile für den hier in Frage stehenden Verwertungszweck - studentische Beurteilung von Studienbereichen - überhaupt bedeutsam?

Dies ist hier in der Tat der Fall. Beziehen wir nämlich die Untergliederung der vormaligen „Mittelgruppe“ auf die Art der Hochschulen (Unis/Gesamthochschulen versus Fachhochschulen), so zeigt sich, dass deren Vernachlässigung einen erheblichen Informationsverlust (nicht lediglich eine Reduktion des Detaillierungsgrades) zur Folge hätte. Die Kritik am Praxisbezug der Ausbildung und an der didaktischen Vermittlung des Stoffs trifft vor allem die Universitäten (der %-Anteil der Kritiker ist hier fast doppelt so hoch wie an

Fachhochschulen). Die negative Benotung des Forschungsbezugs und der Internationalität trifft dagegen fast ausschließlich die Fachhochschulen (der %-Anteil ist hier dreimal so hoch wie an den Universitäten!).

CLLEHR4 Stud. Bewertung	n	F1 Art der Hochschule		Total
		Uni/GH	FH	
		1 alles gut	2555	
	34,1%	32,2%	32,8%	
2 Kritik Praxis	2811	3100	5911	
	37,6%	20,0%	25,7%	
3 Kritik Forschung	759	4803	5562	
	10,1%	30,9%	24,2%	
4 alles schlecht	1360	2624	3984	
	18,2%	16,9%	17,3%	
Gesamt	7485	15528	23013	
	100,0%	100,0%	100,0%	

Tabelle 5: Urteilsdimensionen und Hochschultyp

Bildet man ohne Kenntnis dieses unterschiedlichen Objektbezugs aus den Antworten der Studierenden an Universitäten und Fachhochschulen einen Indexwert, der die vergebenen Noten für die einzelnen Statements zusammenfasst⁸, geht dieser für das Verständnis der studentischen Evaluationen bedeutsame Unterschied nicht nur ersatzlos verloren; er wird im Zuge der Analyse nicht einmal für den Forscher auch nur ansatzweise sichtbar.

Nun geht es bei dieser Tagung aber nicht um allgemeine Fragen der Auswertung von Umfragedaten, sondern um die Möglichkeiten eines Ranking. Lässt sich beispielsweise die übersichtliche Dreiergruppierung der Hochschulen in eine eindeutige Spitzengruppe, eine Gruppe mit eher mittelmäßiger Bewertung und eine Schlussgruppe vornehmen? Natürlich - Klassifikationen sind immer machbar; man braucht nur eine Regel für die Klassierung zu formulieren.

Variablenorientierte Regeln beziehen sich dabei auf die Verteilungseigenschaften der Variablen: Entweder werden die Schnittpunkte so gelegt, dass auf jede Gruppe ein vorgegebener Anteil entfällt; z. B. die mittlere Gruppe enthält 50% der „mittleren Fälle“ (Schnittpunkte Median plus/minus ein Quartil), die Spitzengruppe enthält die 25% „besten“, die Schlussgruppe die 25%

⁸ In der CHE-Auswertung geschieht dies beim „Baustein Studium und Lehre“, wo aus den Antworten der Studierenden zu den zehn Items der Frage 4 ein Gesamtindex „Lehrangebot“ gebildet wird (s. Buhr u.a., 2000, S. 17), der dann dazu dient, eine der Rangreihen der Universitäten und Fachhochschulen zu produzieren - nämlich „Studium und Lehre“, Untertitel: „Wo die Studenten das Lehrangebot am besten bewerten“ (stern und CHE, 2000).

„schlechtesten“ Fälle⁹. Oder die Schnittpunkte werden so gelegt, dass sie die Grenze für „signifikante Abweichungen“ vom arithmetischen Mittel repräsentieren¹⁰. Beides sind im Prinzip willkürliche Entscheidungen; die erste hat den Vorteil, dass sie auch für den Laien gut nachvollziehbar ist; für die zweite spricht, dass sie das statistisch übliche Verfahren ist.

Auch auf der Basis von Clusterdaten (z. B. der Vier-Cluster-Lösung) lassen sich eine eindeutige und nachvollziehbare Klassierungsregeln leicht formulieren. Beispielsweise: Eine eindeutige Zuordnung eines Studienbereichs zu einer der drei Gruppen erfolgt dann, wenn mindestens 50% der Befragten in diesem Studienbereich sich einig sind; d. h. entfallen mindestens 50% in Cluster 1, dann gehört der Studienbereich zur Spitzengruppe; entfallen mindestens 50% in Cluster 4, dann gehört er zur Schlussgruppe; entfallen mindestens 50% auf Cluster 2 oder 3, dann gehört der Bereich zur mittleren Gruppe. Der Rest, für den sich kein Mehrheitsvotum der Befragten ergibt, wird einer Restgruppe “ohne Mehrheitsurteil” zugeordnet. Letzteres ist dann nicht ein “missing value”, also das Fehlen eines gültigen Wertes, sondern eine empirisch gehaltvolle Aussage: In diesen Studienbereichen ist unter den Studierenden das Qualitätsurteil strittig; jeweils relativ große Teile von ihnen kommen mit den dort gegebenen Möglichkeiten für ihr Studium gut, nur mittelmäßig oder schlecht zurecht. Es dürfte sich hier also entweder um einen Bereich mit heterogenen Studienbedingungen oder mit einer heterogenen Studierendenschaft handeln.

Im folgenden wird dies einmal für das Fach Architektur illustriert (ohne dass sich darin eine Präferenz für dieses Fach ausdrückt: es ist im CHE-Studienführer das erste), allerdings mit einer Ergänzung der Klassifikationsregel: Nur in relativ wenigen Hochschulen existiert unter den Studierenden im Fach Architektur eine Mehrheit für die Einordnung in die Spitzengruppe (hier mit A bezeichnet), die Schlußgruppe (C) oder die Mittelgruppe (B); die Heterogenität der studentischen Meinungen ist in den meisten Hochschulen zu groß. Daher wird zusätzlich formuliert: Fallen 40 bis unter 50% der Studierenden in ein Cluster, werden diese Hochschulen als „annähernd“ zur Spitzen-, Mittel- oder Schlussgruppe gehörig eingeordnet (gekennzeichnet mit [A], [B] bzw. [C]).

⁹ Beim CHE-Ranking für die Gruppierung mit “Fakten”-Daten verwendet.

¹⁰ Beim CHE-Ranking für die Gruppierung nach Befragungsdaten verwendet.

Kurzname HS	CHE- Index Studium u. Lehre	CHE- Rang- ordnung	Rang- ordnung nach Cluster	%-Anteil Cluster 4-1 „alles gut“	%-Anteil Cluster 4-2 „Kritik Praxis“	%-Anteil Cluster 4-3 „Kritik Forschung“	%-Anteil Cluster 4-4 „alles schlecht“
AdBK Stuttgart	2,5	1	A	66,7	15,4	7,7	10,3
Uni-GH Kassel	2,5	1	A	56,3	18,8	15,6	9,4
BTU Cottbus	2,6	1	A	52,7	24,2	15,4	7,7
Uni Weimar	3,0	1	[B]	27,2	44,6	8,7	19,6
TU Dresden	3,4	3	[B]	9,8	40,2	14,7	35,3
TU Braunschw.	3,5	3	[C]	12,3	29,2	16,9	41,5
TU Berlin	3,4	2	[C]	17,0	26,4	15,1	41,5
Uni-GH Wupp.	3,6	3	[C]	10,0	12,0	34,0	44,0
TU München	3,6	3	C	12,2	22,0	12,2	53,7
Uni Hannover	3,6	3	C	7,7	21,2	15,4	55,8
HdK Berlin	2,9	1	---	30,6	32,3	19,4	17,7
Uni Stuttgart	3,0	2	---	27,4	29,0	27,4	16,1
Uni Karlsruhe	3,1	2	---	24,7	31,2	24,7	19,4
Uni Kaisersl.	3,3	2	---	21,9	28,8	13,7	35,6
Uni Dortmund	3,1	2	---	21,5	24,6	32,3	21,5
RWTH Aachen	3,2	2	---	19,0	32,8	17,2	31,0
TU Darmstadt	3,3	2	---	18,6	32,9	15,7	32,9
Total	3,2			24,1	28,9	17,8	29,2

Tabelle 6: Klassifikation des Studienbereichs Architektur: Universitäten

Auch nach dieser „weicheren“ Regel bleiben von 17 Universitäten 7 übrig, die nicht eindeutig einzuordnen sind. Für sie kann aufgrund der studentischen Befragungsdaten jede Art von beratender Aussage gerechtfertigt sein: positive Empfehlung, entschiedene Warnung, differenzierende Einschränkung: „ganz gut, aber Praxisbezug und Didaktik lassen zu wünschen übrig“. Beispiel Uni Dortmund: Knapp 22% halten alles für gut, genauso viele sehen alles negativ, fast 25% üben an Praxisbezug und Didaktik Kritik, 32% an Forschungsbezug und Internationalität. Zumindest für solche Studienbereiche wären ergänzende Informationen darüber nötig (und für die Zielgruppe des CHE-Rankings von Interesse), warum das Urteil so strittig ausfällt.

Ähnlich die Klassierung für den Studienbereich Architektur an Fachhochschulen: In 17 von 49 Fachhochschulen ist die Heterogenität der Urteile so groß, dass eine klare Eingruppierung sich verbietet. Nur wenige fallen eindeutig in die Spitzen- oder die Schlussgruppe; und auch die „Randbereiche“ der Spitzen- und der Schlussgruppe (mindestens 40% Lob oder Kritik) bleiben relativ schwach besetzt. Deutlich größer ist hier allerdings die Zahl der Hochschulen, die mehrheitlich bzw. von mindestens 40% der Befragten als „mittel, mit Kritik an Teilaspekten“ beurteilt wird.

Kurzname HS	CHE-Index Studium u. Lehre	CHE- Rang- ordnung	Rang- ordnung nach Cluster	%-Anteil Cluster 4-1 „alles gut“	%-Anteil Cluster 4-2 „Kritik Praxis“	%-Anteil Cluster 4-3 „Kritik Forschung“	%-Anteil Cluster 4-4 „alles schlecht“
FH Biberach	2,4	1	A	66,7	8,3	18,3	6,7
FH Heidelberg	2,4	1	A	59,1	9,1	27,3	4,5
HS Muthesius	2,6	1	A	50,0	28,6	21,4	,0
FH Bochum	2,7	1	[A]	46,5	18,3	25,4	9,9
FH Lausitz	2,8	1	[A]	44,4	3,7	40,7	11,1
FH Konstanz	2,7	1	[A]	40,0	16,4	38,2	5,5
FH Würzburg	3,2	2	B	5,6	1,9	75,9	16,7
TFH Berlin	3,1	2	B	15,5	1,7	60,3	22,4
FH München	3,0	2	B	17,3	10,2	56,1	16,3
FH Lüneburg	3,1	2	B	19,7	8,5	54,9	16,9
FH Weihenst.	3,1	2	B	16,7	13,5	52,1	17,7
HS Wismar	2,8	1	B	37,2	4,7	51,2	7,0
FH Erfurt	3,1	2	B	21,3	11,0	50,3	17,4
FH Kiel	3,4	3	B	7,4	11,1	50,0	31,5
FH Karlsruhe	3,2	2	[B]	22,5	7,5	48,8	21,3
FH Oldenburg	3,1	2	[B]	18,3	12,7	47,9	21,1
FH Darmstadt	2,9	1	[B]	26,7	9,9	47,5	15,8
FH Nürnberg	3,2	2	[B]	10,9	12,7	47,3	29,1
FH Stuttgart	2,9	1	[B]	32,7	10,9	46,4	10,0
FH Lübeck	3,1	2	[B]	21,6	14,9	45,9	17,6
HS Anhalt	3,0	2	[B]	22,7	19,7	45,5	12,1
HTW Dresden	2,9	1	[B]	32,3	9,7	45,2	12,9
FH Hannover	3,3	2	[B]	10,4	18,8	43,8	27,1
FH Augsburg	2,9	1	[B]	33,7	6,7	43,8	15,7
FH Wiesbaden	3,1	2	[B]	19,3	14,3	42,0	24,4
Uni-GH Essen	3,1	2	[B]	12,0	28,0	42,0	18,0
FH Bielefeld	3,1	2	[B]	25,0	16,1	41,1	17,9
FH Mainz	3,4	3	[C]	14,1	8,9	36,3	40,7
HS Bremen	3,6	3	[C]	13,3	8,9	28,9	48,9
FH Düsseldorf	3,8	3	C	1,0	12,4	27,8	58,8
FH Hamburg	4,0	3	C	5,6	0,0	22,2	72,2
HTWK Leipzig	2,9	1	---	37,5	15,6	37,5	9,4
FH Hildesheim	2,8	1	---	36,6	17,9	36,6	8,9
FH Potsdam	2,9	1	---	36,1	25,0	22,2	16,7
FH Koblenz	3,1	2	---	27,5	17,5	35,0	20,0
FH Regensburg	3,2	2	---	23,0	11,5	39,3	26,2
FH Coburg	3,2	2	---	20,9	11,6	37,2	30,2
FH Dortmund	3,4	3	---	20,8	6,3	33,3	39,6
HTW Saarbrück.	3,3	2	---	17,4	21,7	26,1	34,8
Uni-GH Siegen	3,4	2	---	17,4	15,2	39,1	28,3
Uni-GH Paderb.	3,4	3	---	16,4	10,4	34,3	38,8
FH Münster	3,3	2	---	14,3	12,7	39,7	33,3
FH Rosenheim	3,4	3	---	13,9	19,4	37,5	29,2
FH Lippe	3,4	3	---	12,6	19,4	35,9	32,0
FH Kaisersl.	3,5	3	---	11,3	18,1	31,9	38,8
FH Frankfurt	3,3	2	---	11,1	22,2	28,9	37,8
FH Köln	3,3	2	---	5,9	20,6	35,3	38,2
FH Trier	3,4	3	---	5,3	22,1	37,9	34,7
Total	3,2			21,4	13,5	40,8	24,3

Tabelle 7: Klassifikation des Studienbereichs Architektur: Fachhochschulen

Erinnern wir uns an Tab. 5, aus der hervorging, dass die Kritik an Praxis und Didaktik an Universitäten häufiger, an Fachhochschulen selten auftrat; das Umgekehrte galt für die Kritikdimension Forschung/Internationalität. Ein nochmaliger Blick auf die Rangordnungstabellen (Tab. 6 und 7) zeigt nun, in welchem Maße diese vorgenommene Differenzierung des Qualitätsurteils „mittel“ klassifikationsrelevant geworden ist: Bei den Universitäten ist in der Mittelgruppe nur die Kritik an Praxisbezug und didaktischer Vermittlung mit Anteilen von 40% und mehr vertreten. In der Mittelgruppe der Fachhochschulen gilt dies dagegen für keinen einzigen Fall. Alle hier eingeordneten Hochschulen finden sich vielmehr deshalb hier wieder, weil aus studentischer Sicht Forschungsbezug und Internationalität schlechte Noten verdienen.

Wie fällt nun der Vergleich des hier (auf dem Weg über eine Klassifikation der „Evaluatoren“) konstruierten Rankings mit dem von CHE (auf der Basis eines zusammenfassenden Indexes) gewonnenen Resultat aus?¹¹ Wie zu erwarten, gibt es Übereinstimmungen und Abweichungen (jede Klassifikationsregel setzt eine spezifische „Ähnlichkeits“-Definition und impliziert damit auch je spezifische Gruppierungsergebnisse):

- Alle Studienbereiche, in denen mindestens 40% der Befragten das Evaluationsprofil „alles gut“ vertreten (die also nach der oben formulierten Klassifikationsregel zur „Spitzengruppe“ gezählt werden), finden sich auch im CHE-Ranking in der „Spitzengruppe“. Das gleiche gilt analog für die Zugehörigkeit zur „Schlussgruppe“.

Erstes Fazit: Bei relativ homogenem Befragten-Urteil ist das Klassifikationsresultat methodenunabhängig.

- Ein relativ großer Teil der von CHE nach der Regel „signifikante Abweichung vom Index-Mittelwert“ zur Spitzen- oder zur Schlussgruppe zählenden Fälle gelangen nach dem hier verwendeten Verfahren dagegen entweder in die Mittelgruppe oder erweisen sich als Studienbereiche, die in der subjektiven Studierendensicht als so vielfältig erscheinen, dass sich keine Mehrheit (nicht einmal ein „qualifizierter Anteil“ von 40%) für ein klares Urteil findet.

Zweites Fazit: Die Regel „Die Mehrheit (oder ein qualifizierter Anteil) vertritt ein interpretierbares Urteilsprofil“ führt nicht nur zu einem inhaltlich nachvollziehbareren (interpretierbareren) Resultat, sondern stellt auch schärfere Anforderungen für die Zuordnung zur Spitzen- oder zur Schlussgruppe als der Bezug auf die künstlich geschaffene Reduktion der mehrdimensionalen Evaluationen zu einem eindimensionalen Konstrukt „Qualität von Studium und Lehre“: Wird die interne Widersprüchlichkeit von Evaluationsurteilen „weggemittelt“, genügt bereits eine „signifikante“ (und das ist bei großer Fallzahl auch eine faktisch sehr geringe) statistische Tendenz, um scheinbar eindeutig als „Spitze“ zu gelten (oder zu den „Schlusslichtern“ zu gehören).

- Bereits die Reduktion auf ein eindimensionales Evaluationskonstrukt „Qualität von Studium und Lehre“ führt - darauf wurde oben mehrfach hingewiesen - zu unnötigem (und für den hier verfolgten Zweck unzulässigem) Verzicht auf einen wesentlichen Teil der mit großem Erhebungsaufwand gewonnenen Befragungsinformationen: Auch (besser: gerade) die Heterogenität von Wahrnehmungen bei objektiv identischem Sachverhalt ist von hohem

¹¹ In Tab. 6 und 7 sind in den Spalten 2 und 3 der CHE-Indexwert und die CHE-Klassifikation wiedergegeben.

Informationswert. Die für das Ranking von CHE getroffene Entscheidung, die Mittelgruppe als nicht interpretierbare Restkategorie zu konzipieren,¹² erhöht diesen freiwilligen Informationsverzicht.

Drittes Fazit: Statt die größte Gruppe von Hochschulen (rund 50% der Fälle) als nicht klar gruppierbaren „Rest“ zu behandeln, über den lediglich ausgesagt wird, dass er nicht eindeutig zur „Spitze“ und auch nicht eindeutig zum „Schluss“ gehört, erscheint es rational sinnvoller, zusätzliche Informationen aus den Befragungsdaten über diesen diffusen „Rest“ herauszuarbeiten. Immerhin ist es für die Zielgruppe des Studienführers nicht uninteressant zu erfahren, ob die Befragten den Bereich, in dem sie studieren, in der Tat als mittelmäßig oder „durchschnittlich“ ansehen, oder ob dort sowohl gute wie schlechte Noten von jeweils relativ vielen Studierenden vergeben werden.¹³ So könnte z. B. eine zur „tatsächlichen Mittelgruppe“ gehörende Hochschule für solche Studieninteressierten der Ort ihrer Wahl sein, die kein hohes Anspruchsniveau mitbringen (etwa weil sie lediglich aus Gründen der beruflichen Karriere „irgend einen“ akademischen Abschluss benötigen). Andererseits könnte sich auch eine Hochschule mit heterogener Beurteilung für Studieninteressierte durchaus als „Spitze“ erweisen - nämlich dann, wenn sie zur selben Zielgruppe gehören wie diejenigen, die dort das Urteil „alles gut“ gefällt haben.

6. Fehler bei der Interpretation der Daten sowie bei Schlussfolgerungen und Empfehlungen

Die im vorigen Abschnitt angeführten Beispiele geben mir eine gute Überleitung zu der zu Beginn angesprochenen dritten Fehlerquelle: **‚Fehler bei der Interpretation der Daten sowie bei Schlussfolgerungen und Empfehlungen‘**.

Im Rahmen dieser Veranstaltung will ich mich allerdings auf den ersten Aspekt beschränken; die Frage der aus Hochschulrankings zu gewinnenden Schlussfolgerungen und Empfehlungen überlasse ich anderen; ebenso Antwort auf die Frage: Lohnt sich der ganze Aufwand?

Natürlich gibt es Überschneidungen zwischen „Interpretationen“ und „Schlussfolgerungen“; denn es versteht sich wohl von selbst, dass alle diejenigen Schlussfolgerungen notwendigerweise falsch sind, die sich zu ihrer Begründung auf statistische Artefakte berufen. Wenn schon die empirischen Daten falsch interpretiert werden, können auch die praktischen Ableitungen daraus nicht zutreffend sein. Darüber hinaus müssen Fehlschlüsse und in die Irre führende Empfehlungen aber auch dann eintreten, wenn die Daten zwar einen Sachverhalt korrekt abbilden (also keine Artefakte sind), aber uninterpretiert bleiben: Was bedeutet ein „Mehr“ an Absolventen? Wäre es nicht vernünftig, manchen Studierenden an der Universität zu raten, von der wissenschaftlichen zur Fachhochschul-Ausbildung zu wechseln (auch wenn durch solche Beratung die „Abbrecherquote“ erhöht würde)? Was folgt aus einer rigiden Verkürzung der

¹² Nur die Mittelwertabweichungen zwischen den Zugehörigen zur Spitzen- und zur Schlussgruppe gelten als „statistisch signifikant“ (Buhr u.a., 2000: S. 23; start 1/2000: S. 38 ff.)

¹³ Entweder weil die Studiensituation selbst durch Vielfalt und/oder Unterschiedlichkeit geprägt ist oder weil in diesen Bereichen eine überdurchschnittlich heterogene Studierendenschaft anzutreffen ist.

Studienzeit? Ist damit wirklich ein Qualitätsgewinn verbunden? Was sagt ein niedrigeres Durchschnittsalter der Absolventen aus? Was bedeutet eine Erhöhung des Anteils ausländischer Studierender? Sind 50% „viel besser“ als 5%? All dies kann nicht Selbstzweck sein; und all dies ist nicht per se ein Merkmal von Qualität.

Wovor ich also warnen möchte, ist die häufig zu beobachtende Praxis, bei den reinen empirisch ermittelten Zahlenwerten stehen zu bleiben, sie als „Fakten“ zu betrachten, die „für sich selbst sprechen“. Alle mit empirischen Methoden zu ermittelnden Informationen sind allenfalls Indikatoren für komplexere Sachverhalte (manchmal nicht mal dieses). Sie müssen in Vorstellungen über die Realität zurückübersetzt, sie müssen „interpretiert“ werden. Wenn sich empirische Indikatoren zu automatisiert ablaufenden Entscheidungsmechanismen verselbständigen, ist es ein Leichtes, die relevanten Indikatoren zu optimieren, sogar so weit, dass der größere Kontext darunter leidet, der durch den Indikator repräsentiert werden sollte.¹⁴

Zur Illustration will ich noch einmal auf das Thema Praxisbezug und Forschungsbezug zurückkommen. Ich hatte gezeigt, dass von den Studierenden an den Universitäten der Praxisbezug, von den Studierenden an den Fachhochschulen der Forschungsbezug schlechter benotet wird. Was kann daraus geschlossen werden? Sind die Hochschulen - jeder Typ auf seine Art - defizitär? Sind daran eventuell sogar Empfehlungen nach dem einfachen (aber beliebten) Strickmuster zu knüpfen?: „Dies wird als Defizit wahrgenommen, also muss es beseitigt werden; da müssen eben die Universitäten ihre Lehre stärker auf die Praxis ausrichten und die Fachhochschulen die Forschung etwas stärker in den Blick nehmen.“ Wenn eine solche Empfehlung angenommen und mit Erfolg umgesetzt würde - welche Konsequenzen hätte das für unser Hochschulsystem?

Man könnte jedoch auch etwas ganz anderes daraus schließen: Die betreffenden Studierenden haben wohl den falschen Hochschultyp gewählt. Die mehr Praxis fordernden Universitätsstudenten gehören eigentlich an die Fachhochschule, die mehr Forschung fordernden Fachhochschulstudenten sollten sich einen Wechsel an die Universität überlegen.

Und man kann schließlich auch ein Drittes folgern, nämlich: Da ist die falsche Frage gestellt worden. Ein Beispiel: Stellen Sie sich vor, man ließe von Verbrauchern Getränke testen und bewerten, u. a. Mineralwasser und Cola light; und ein Kriterium, das abgefragt wird, sei die Süße des Getränks. Unter diesem Kriterium wird Cola light sicher als „süß“ bewertet, Mineralwasser wohl eher nicht (sofern nicht einige der Tester dieses Item für Mineralwasser als unangemessen zurückweisen). Ist nun Cola light „besser“ als Mineralwasser, weil „süßer“? Ja - wenn wir für alle Getränke das gleiche Bewertungsraster anlegen. Bei Fachhochschulen und Universitäten geschieht genau dies, wenn aus den Antworten der Studierenden auf die in beiden Hochschultypen gleiche Fragenbatterie ein Gesamtindex gebildet wird.

Haben die Studierenden das so gewollt? Das ist zu bezweifeln. Von den Fachhochschul-Studierenden, die anmerken, dass es an ihrer Hochschule mit Forschungsbezug und Internationalität nicht weit her sei, charakterisieren immerhin zwei Drittel dieses Faktum als für sie persönlich wenig wichtig bis völlig unwichtig. Der Befund ist also wie folgt zu lesen: „Das

¹⁴ Auch auf diese Frage muss ich hier nicht im Detail eingehen und kann auf den o.g. Aufsatz von Christian Bayer verweisen.

Mineralwasser ist zwar nicht süß; aber das macht nichts.“ Das gilt übrigens sowohl für den hier speziell betrachteten Bereich Architektur als auch über alle Studienbereiche hinweg:

F1 Art der Hoch- schule	CLLEHR4 Cluster Stud. Bewertung	F13#1 Wichtig Praxisbezug		F13#2 Wichtig Forschungsbezug	
		mittlerer Skalenwert	% Skalenwert 1 od. 2	mittlerer Skalenwert	% Skalenwert 1 od. 2
		Uni/GH	1 - alles gut	1,6	89,0
	2 - Kritik Praxis	1,7	86,8	2,5	53,0
	3 - Kritik Forschung	1,7	87,8	2,7	45,7
	4 - alles schlecht	1,6	89,1	2,7	45,9
	Total	1,7	88,1	2,5	56,5
FH	1 - alles gut	1,3	98,1	2,6	49,3
	2 - Kritik Praxis	1,4	94,8	2,7	45,2
	3 - Kritik Forschung	1,4	96,3	3,0	34,6
	4 - alles schlecht	1,4	94,3	2,9	40,9
	Total	1,4	96,3	2,8	42,5

Tab. 8: Wichtigkeit von Praxis- und Forschungsbezug (alle Studienbereiche)

Forschungsbezug und Internationalität werden also an Fachhochschulen nicht nur mit schlechteren Benotungen belegt als an Universitäten (s. o.); sie gelten den dortigen Studierenden auch als nicht so wichtig. Und das gilt in besonderem Maße für diejenigen Studierenden, die auf die entsprechenden Fragen schlechte Noten vergeben haben! Das wahrgenommene Fehlen dieser Eigenschaften wird also wohl auch nicht als Qualitätsmangel (sondern eher als „normales“ Charakteristikum des gewählten Hochschultyps) empfunden. Dies sollte bei der Interpretation des obigen Rankings für die zur „tatsächlichen Mittelgruppe“ gehörenden Fachhochschulen berücksichtigt werden!

Umgekehrt werden Praxisbezug und Vermittlung des Lernstoffs an Fachhochschulen positiver beurteilt als an Universitäten. Und im Unterschied zum Forschungsbezug ist diese Qualitätsdimension fast allen Studierenden sehr wichtig oder wichtig (Skalenwerte 1 und 2), und dies weitgehend unabhängig vom Hochschultyp (96% an Fachhochschulen und 88% an Universitäten kreuzen die Skalenwerte 1 oder 2 an!) und ebenfalls unabhängig vom Beurteilungsprofil der Befragten.

7. Zur Validität von Studierendenbefragungen als Evaluation

Abschließend ein kurzes Resümee als Antwort auf die Frage nach der Validität von Studierendenbefragungen als Evaluation und zu ihrem Informationsgehalt.

Selbstverständlich liefern Befragungen von Studierenden so valide Resultate wie die Befragung anderer Gruppen der Bevölkerung. Selbstverständlich sind Studierende in der Lage, subjektiv zutreffende Urteile über ihre Situation des Studiums und über ihre Wahrnehmung der Lehre zu fällen - so wie jede Klientengruppe, die eine Dienstleistung in Anspruch nimmt, diese Dienstleistung beurteilen kann. Aber: Solche Beurteilungen sind keine objektivierbaren Evaluationen der Sachverhalte, auf die sich die Bewertungsfragen richten. Und sie sollen dies in anderen Klienten- und Kundenbefragungen auch gar nicht sein. Selbstverständlich wird etwa der Hersteller eines Automobils nicht die Kunden befragen, wie denn die Qualität seines Produkts sei (das testet er auf andere und angemessenere Weise). Vielmehr will er etwas über seine Zielgruppe wissen, für die er das Produkt herstellt und an die er es verkaufen will. Er will nicht Evaluation, sondern Zielgruppenanalyse und Akzeptanzinformationen. Und darüber geben Klientenbefragungen (in den generellen Grenzen der Validität von Befragungen) valide Auskunft. Akzeptanz ist im übrigen für die Produktion von Dienstleistungsqualität eine wichtige Voraussetzung .

Damit komme ich zum Informationsgehalt solcher Befragungen. Er ist potentiell als sehr hoch einzuschätzen (wird aber häufig nicht hinreichend ausgeschöpft). Wie anders als über die subjektiven Urteile der Betroffenen sollte man über deren Verhältnis zur angebotenen Dienstleistung etwas erfahren? Da aber dieses Verhältnis von einer Reihe von Merkmalen der Befragten und ihrer individuellen Situation abhängt, benötigt man Zusatzinformationen, damit die erhobenen Einschätzungen und Urteile für den auswertenden Forscher überhaupt einen rekonstruierbaren Sinn erhalten. Eine isolierte Erhebung nur der Beurteilungen selbst ist ziemlich informationsleer und führt zwangsläufig in die Gefahr übervereinfachender und häufig falscher Schlussfolgerungen und Empfehlungen.

Die Frage nach Validität und Informationsgehalt studentischer Evaluationsbefragungen ist also umzuformulieren. Sie sollte nicht heißen: Wie valide und wie informativ sind studentische Bewertungen? Sondern: Wie valide und wie informationsausschöpfend werden studentische Bewertungen erhoben, ausgewertet und verwertet?

Literatur

Bayer, Christian R., 2000, „Anspruch und Wirklichkeit von Hochschul-Ranking: Vorschlag einer allgemeinen Methodik“. *Die Betriebswirtschaft* 5/2000.

Bayer, Christian R s. auch den Beitrag des Autors in diesem Tagungsband, S. 91 ff.

Buhr, Petra; Giebisch, Petra; Hornbostel, Stefan; Müller-Böling, Detlef. „Der Studienführer 2000. Vorgehensweise und Indikatoren“. *CHE-Arbeitspapier Nr. 22*. Gütersloh 2000

Informationen Bildung Wissenschaft, Nr. 7-8/1992

MWF NW. *Aktionsprogramm Qualität der Lehre. Abschlussbericht*. Düsseldorf 1991

Schmidt, Jörn. *Evaluation I. Evaluation als Diagnose. HDZ-Dozentenkurs, (Hochschuldidaktisches Zentrum)*. Essen 1980

stern und CHE. *start. Der Studienführer. Erste Hilfe für das Studium*. Hamburg 2000

Bootstrap-Rangkonfidenzintervalle für Hochschul-Ranglisten

Uwe Jensen¹

Zusammenfassung

Dieser Vortrag befasst sich mit den vorhandenen Möglichkeiten, die Lehrleistung von Hochschulen zu messen, und zeigt, warum trotz aller Vorbehalte auf die Evaluation durch Studentenbefragung nicht verzichtet werden kann. Danach wird exemplarisch die Vorgehensweise des SPIEGELS bei der Erstellung seiner Hochschulranglisten kritisch analysiert. Dabei werden fünf Aspekte herausgearbeitet, die gegenwärtig nicht im Zentrum der Debatten stehen, aber deutlich verbesserungswürdig sind. Als kritisch erweist sich insbesondere die ungenaue Quantifizierung der Genauigkeit von Ranglisten. Daher wird ein neues Bootstrap-Verfahren zur Erstellung von Rangkonfidenzintervallen vorgestellt und in Simulationen untersucht. Abschließend erfolgt die Anwendung des Verfahrens auf einen aktuellen Datensatz des CHE.

1. Einleitung

Ein bedeutender Teil der öffentlichen Debatte um die Hochschulpolitik kreist um die Möglichkeit und Durchführung der Messung von Hochschulleistungen in Forschung und Lehre. Es wird gefordert, Professorengehälter nach Leistung zu zahlen und finanzielle Mittel an Universitäten nach Leistung zu vergeben. Es ergibt sich daher die Frage, wie denn Hochschulleistung zu messen ist. Bei den von privaten Unternehmen verkauften Produkten und Dienstleistungen erfolgt diese Bewertung über den Markt. Bei den öffentlichen Hochschulen entfällt diese automatische Bewertung größtenteils, sodass hier ein System von Leistungsindikatoren geschaffen werden muss.

Seit den 80er Jahren befassen sich zahlreiche Veröffentlichungen mit dem Problem der Schaffung von geeigneten Indikatorensystemen, so z. B. Daniel und Fisch (1988), Fisch und Daniel (1986), Helberger (1989 und 1991) oder Schenker-Wicki (1996). Aber während in den anglo-amerikanischen Ländern diese Instrumente seit den 80er Jahren auch in der Praxis eingesetzt werden, scheute man davor in Deutschland zurück. Für den Bereich Lehre übernahm der SPIEGEL die Vorreiterrolle durch seine 1990 erschienene und sogleich heftig diskutierte und kritisierte Hochschulrangliste. Weitere durch die Medien organisierte Ranglisten folgten, z. B. durch FOCUS 1997, durch den SPIEGEL 1993 und 1999 und die Studien durch STERN/START/CHE von 1999 und 2000. Das ‚Centrum für Hochschulentwicklung (CHE)‘, eine von der Hochschulrektorenkonferenz und der Bertelsmann-Stiftung getragene Einrichtung, nimmt dabei eine bedeutende Rolle ein.

All diese Ranglisten wurden und werden weiterhin von teilweise heftiger Kritik begleitet (wie etwa in Meinefeld (2000)), es scheint aber jedenfalls im Bereich Lehrleistung keine erfolgreiche

¹ Ich danke Gerd Hansen und Stefan Mittnik sehr für ihre Unterstützung. Stefan Hornbostel gilt mein Dank für die Zusendung eines Vorabdrucks einer Veröffentlichung und eines CHE-Datensatzes. Hilfreiche Kommentare verdanke ich zahlreichen Mitgliedern der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Kiel sowie einigen Teilnehmern der CHE-Tagung zu Hochschulrankings in Berlin und der Statistischen Woche in Nürnberg. Alle verbliebenen Fehler habe ich zu verantworten.

wissenschaftliche Diskussion darüber stattzufinden, wie denn bessere Hochschulleistungsmessung durchzuführen wäre (vgl. Hornbostel (2000: 1)).

Abschnitt 2 dieser Arbeit befasst sich mit den vorhandenen Möglichkeiten, die Lehrleistung von Hochschulen zu messen, und zeigt, warum auf die Evaluation durch Studentenbefragung nicht verzichtet werden kann. Der folgende Abschnitt skizziert die Vorgehensweise bei der Erstellung der SPIEGEL-Rangliste, um für die nachfolgende Diskussion einen Referenzpunkt zu haben. Abschnitt 4 nennt 10 häufig vorgebrachte und/oder wichtige Kritikpunkte an der Vorgehensweise des SPIEGELS. Es zeigt sich, dass zahlreiche heftige Debatten über die Qualität von derartigen Ranglisten auf den falschen Schauplätzen stattfinden. Das relativ akzeptierte Ranggruppenverfahren, das der Ungenauigkeit von Ranglisten Ausdruck verschaffen soll, ist selbst viel zu ungenau und sollte daher durch ein im sich anschließenden Abschnitt vorgestelltes Bootstrap-Verfahren zur Erstellung von Rangkonfidenzintervallen ersetzt werden. Abschnitt 6 untersucht mit Hilfe einer Simulation die Eigenschaften der beiden Verfahren. Abschließend werden Bootstrap-Verfahren auf einen aktuellen Datensatz des CHE angewendet.

2. Hochschulleistungsindikatoren

Hochschulen produzieren in Verbundproduktion (siehe z. B. Backes-Gellner (1989a)) Forschung und Lehre. Diese Arbeit hat die Messung der Lehrleistung zum Thema. Die Messung der Forschungsleistung wird wegen einiger methodischer Unterschiede nicht betrachtet. Leistungsprüfungen können u. a. auf die Überprüfung von Effizienz (als Relation von Output zu Input) oder Effektivität abzielen (Schenker-Wicki (1996: 11)). Die Effizienzmessung für Hochschulleistungen, mit der sich etwa Backes-Gellner (1989b) befasst, soll im folgenden ausgeklammert werden.

Welche Leistungsindikatoren gibt es für den Bereich Hochschullehre? Schenker-Wicki (1996: 114 ff) u. a. nennen z. B.: Anzahl Studenten, Bewerber pro Studienplatz, Betreuungsverhältnisse, Durchschnittsalter der Absolventen, durchschnittliche Studiendauern, Diplom-, Promotions-, Habilitations-Quoten, Examenserfolgsquoten, Abbruchquoten, Durchschnittsnoten, Beschäftigungsgrad, Netto-Einkommensertrag nach n Jahren, Anzahl ausländischer Studierender, Anzahl Studierender mit Fremdsemestern, Anzahl Gastprofessoren.

Die zahlreichen Nachteile dieser Indikatoren sind seit vielen Jahren bekannt (vgl. etwa Schenker-Wicki (1996: 117 ff)). Einige Punkte lauten: Die Anzahl der Studenten und die Anzahl der Bewerber pro Studienplatz messen nur Quantität, aber nicht Qualität und hängen zudem oft auch von der Größe des Einzugsgebietes und natürlich dem Studienfach ab. Betreuungsverhältnisse spiegeln die finanzielle Ausstattung wider, um deren gerechte Ermittlung es bei der Hochschulleistungsmessung u. a. geht. Ähnliches trifft teilweise für die Indikatoren Durchschnittsalter der Absolventen und durchschnittliche Studiendauern zu. Diplom-, Promotions-, Habilitations-Quoten, Examenserfolgsquoten und Durchschnittsnoten wären (sollten diese zu wesentlichen Lehrleistungsindikatoren erkoren werden) in für die Ausbildungsqualität schädlicher Weise manipulierbar. Hohe Abbruchquoten können ein Indiz für schlechte Lehre, aber auch für schlechte finanzielle Ausstattung oder konsequente Elimination ungeeigneter Studenten sein. Der Beschäftigungsgrad n Jahre nach dem Abschluss hängt auch von außeruniversitären Faktoren ab. Der Netto-Einkommensertrag der

Hochschulbildung (siehe Tracy und Waldfogel (1997)) ist u. a. wegen der bei der Einkommensfunktionsschätzung bekannten Probleme der Messung individueller Fähigkeiten (siehe Jensen (2001)) schwer quantifizierbar.

Kein Indikator wird also der Komplexität des Produkts ‚Hochschullehre‘ gerecht, kein Aggregat dieser Indikatoren ist hinreichend allgemein einsetzbar oder gar akzeptiert. Daher kommt der Evaluation durch Befragung der Konsumenten der Lehrleistung, den Studenten, eine große Bedeutung zu. Auch dieses Prüfverfahren hat natürlich Nachteile, die im übernächsten Abschnitt diskutiert werden. Aber schon die Tatsache, dass in den Medien regelmäßig auf diese Weise die Lehrleistung von Hochschulen bewertet wird, macht eine wissenschaftliche konstruktive Analyse der Vor- und Nachteile dringend erforderlich.

3. Die SPIEGEL-Ranglisten

Der SPIEGEL hat mit seinen Ranglisten von 1990, 1993 und 1999 den Ausgangspunkt für alle weiteren Studien in diesem Bereich geschaffen. Daher sollen anhand der SPIEGEL-Rangliste des Jahres 1993, bei der schon einige Anfangsmängel der ersten Rangliste beseitigt worden sind, die generelle Vorgehensweise und die möglichen Nachteile betrachtet werden.

Im Januar/Februar 1993 wurden in $U^* = 57$ westdeutschen (und 12 ostdeutschen) Universitäten Hauptfachstudenten im 5. bis 10. Semester, die schon mindestens ein Semester an der zu beschreibenden Universität studiert hatten und im gegenwärtigen Semester mindestens 4 Semesterwochenstunden belegt hatten, befragt. Berücksichtigt wurden nur die 15 meiststudierten Fächer oder Fachgruppen (darunter Ökonomie und Politik/Sozialwissenschaften), die von etwa 70 % aller Studenten belegt werden. Ausgeschlossen von der Untersuchung waren Fachhochschulen, die Universitäten der Bundeswehr, die Fern-Universität Hagen, die Katholische Universität Eichstätt und private Universitäten. Ferner wurden nur Fachbereiche mit mindestens 200 Studenten und (für die Gesamtbeurteilung) Universitäten mit mindestens 4 Fakultäten aufgenommen (SPIEGEL (1993: 8f und p. 153)).

Durch mindestens 2 verschiedene Interviewer wurden pro Fachbereich und Universität an mindestens drei verschiedenen Standorten $S = 18$ Studenten zufällig ausgewählt und befragt. Diese mussten u. a. $N = 18$ Fragen $X_{s,n}$ (mit $n = 1, \dots, N$ und $s = 1, \dots, S$) zur Lehrqualität mit den Noten 1 (für den Idealfall) bis 6 (für das negative Extrem) beantworten. Die arithmetischen Mittel dieser $S \cdot N = 324$ Einzelnoten wurden dann über die studentischen Urteile

$$\bar{X}_{s,F,u} = \frac{1}{N} \sum_{n=1}^N X_{s,n} \quad (1)$$

als die Fachbereichsnoten der jeweiligen Universität

$$\bar{X}_{F,u} = \frac{1}{S} \sum_{s=1}^S \bar{X}_{s,F,u} \quad (2)$$

mit $u \in \{1, \dots, U^*\}$ verwendet. Außerdem wurden die zugehörigen Standardabweichungen

$$Std(X)_{F,u} = \sqrt{\frac{1}{S \cdot N} \sum_{s=1}^S \sum_{n=1}^N (X_{s,n} - \bar{X}_{F,u})^2} \quad (3)$$

sowie das Gesamtfachmittel

$$\bar{X}_F = \frac{1}{U} \sum_{u=1}^U \bar{X}_{F,u} \quad (4)$$

als das arithmetische Mittel der Fachbereichsnoten über alle $U \leq U^*$ in diesem Fach berücksichtigten Universitäten berechnet (SPIEGEL (1993: 9)).

Mittels der Fachbereichsnoten wurde eine Fachrangliste der Universitäten erstellt. In dieser Fachrangliste wurden, um der Ungenauigkeit der Rangliste Rechnung zu tragen, die Universitäten im sogenannten Ranggruppenverfahren drei Ranggruppen zugeordnet (RaGruZu): Fachbereiche, deren 90 %-Kondidenzintervall

$$\bar{X}_{F,u} \pm 1,645 \cdot Std(X)_{F,u} \quad (5)$$

vollständig besser als das Gesamtfachmittel \bar{X}_F ist, bei denen also

$$\bar{X}_{F,u} \pm 1,645 \cdot Std(X)_{F,u} < \bar{X}_F \quad (6)$$

gilt, werden in der Spitzengruppe eingeordnet. Ist das 90%-Kondidenzintervall (5) des Fachbereichs vollständig schlechter als das Gesamtfachmittel \bar{X}_F , gilt also

$$\bar{X}_{F,u} \pm 1,645 \cdot Std(X)_{F,u} > \bar{X}_F, \quad (7)$$

so wird der Fachbereich in die Gruppe der Schlusslichter eingeordnet. Alle übrigen Fachbereiche, deren 90%-Konfidenzintervalle also das Gesamtfachmittel \bar{X}_F enthalten, werden in das Mittelfeld eingeordnet (Daniel und Hornbostel (1993: 170)).

Schließlich erhält jede Universität für jeden Platz in der Spitzengruppe einer Fachrangliste 2 Punkte, für jeden Platz im Mittelfeld 1 Punkt und für Plätze in der Gruppe der Schlusslichter 0 Punkte. Diese Punkte werden für die Universitäten aufsummiert und durch die maximal erreichbare Punktzahl (bedingt durch die Anzahl der Fachbereiche in der Studie) dividiert. Anhand der so ermittelten Prozentzahlen werden die Universitäten in eine Rangliste gebracht (SPIEGEL (1993: 10)).

4. Kritik der SPIEGEL-Ranglisten

In diesem Abschnitt werden die möglichen Nachteile der Vorgehensweise der SPIEGEL-Rangliste betrachtet und die Konsequenzen, die daraus durch nachfolgende Studien bereits gezogen wurden oder zukünftig zu ziehen sind, erörtert.

1. Ist die Auswahl der Beurteilenden so durchgeführt worden, dass sich stabile Resultate unabhängig von den ausgewählten Studenten ergeben (vgl. z.B. Schenker-Wicki (1996: 107) oder Dichev (1999))?

Angeichts der im vorigen Abschnitt skizzierten Vorgehensweise und in Anbetracht der Tatsache, dass die Befragungen von geschulten Mitarbeitern (im Fall der SPIEGEL (1993)-

Rangliste des EMNID-Instituts) durchgeführt wurden, kann man hoffen, dass - bei hinreichend großer Anzahl befragter Studenten pro Universität und Fachbereich (siehe Punkt 9) - eine hinreichend gute Auswahl der Studenten getroffen wurde. Die durchgeführten Überprüfungen der Repräsentativität der Stichprobe wie etwa im SPIEGEL (1993: 153) scheinen diese Vermutung zu bestätigen. Hornbostel (2000: 4f) weist darauf hin, dass sehr unterschiedliche Befunde verschiedener Studien meist auf Unterschiede in den Designs zurückzuführen sind.

2. Sind Studenturteile relevanter als geeignete Indikatoren (Schenker-Wicki (1996: 107f))? Wie im vorigen Abschnitt skizziert, weisen alle bisher bekannten Indikatoren gravierende Mängel auf. Es ist sicherlich vernünftig, wie in der Studie von STERN/START/CHE Studenturteile durch von den Fachbereichen und Studentenwerken direkt erfragte Indikatoren sowie durch Professorenurteile zu ergänzen. Aber sich nur auf diese Indikatoren zu stützen hieße wesentliche Informationen über die Lehrqualität zu vernachlässigen.
3. Sind die studentischen Urteile verzerrt durch nicht berücksichtigte Hintergrundvariablen wie Geschlecht, Alter, Studienengagement (gemessen durch z. B. die Anzahl der Semesterwochenstunden), Leistungsniveau (gemessen durch Durchschnittsnoten, Hilfskraft-Tätigkeit), etc. (Hornbostel (2000: 5))?

Fast alle Untersuchungen hinsichtlich dieser Vermutung ergeben keine Hinweise auf systematische und stabile Verzerrungen studentischer Urteile durch Hintergrundvariablen - vgl. u. a. Hornbostel (2000: 5 ff) und Hornbostel und Daniel (1996). Und wenn schwache Einflüsse auf das studentische Urteil feststellbar sind, so sind diese immer fachspezifisch und nicht universitätsspezifisch (Hornbostel (2000: 6)). Dieses spricht für die Beschränkung der Vergleiche auf Fachbereiche und nicht auf ganze Universitäten siehe Punkt 7.

Bei der Vorgehensweise des CHE (s. STERN/START/CHE (2000: 41)) ergeben sich aber u. U. keine Zufallsstichproben, da dort Fragebögen an Studenten verschickt werden, die natürlich nur auf freiwilliger Basis zurückgesendet werden. Hier kann sich eine Verzerrung der Ergebnisse durch die Auskunftswilligkeit ergeben, die durch die oben erwähnten Untersuchungen mangels Informationen über die Nicht-Auskunftswilligen nicht festgestellt werden kann, von der man also nur hoffen kann, dass sie die studentischen Urteile verschiedener Universitäten in gleicher Weise beeinflusst, also die Ranglisten nicht verzerrt.

4. Sind die Studenten kompetente Urteilsgeber? Messen derartige Studien statt der Qualität der Lehre nur den Grad des ‚Wohlgefühls‘ der Studenten? Fehlen den Studenten die Vergleichsmöglichkeiten? Ist die Kritikbereitschaft von Studenten in Metropolen höher als die von Studenten in Kleinstädten? Diese häufig vorgebrachte Kritik findet sich u. a. in Scheuch (1990), Meincke (1999) oder Meinefeld (2000). Zur Unterstützung dieser Hypothesen dient u.a. die Tatsache, dass große Universitäten signifikant schlechter abschneiden als kleine.

Zur Entkräftung dieser Hypothese führt Hornbostel (2000: 9 ff) einen Vergleich mit den Urteilen der Professoren zu den gleichen Fragen an, der in der STERN/START/CHE-Studie (1999) durchgeführt wurde, da Professoren sicher schon wegen der größeren Vergleichsmöglichkeiten durch ihre mehrfachen Ortswechsel im Laufe ihrer Karriere auch

von diesen Kritikern eine größere Kompetenz zugebilligt wird. Professoren und Studenten stimmen danach aber plausiblerweise in vielen Beurteilungen der Lehrsituation überein. Insbesondere typische Problemlagen an großen Fachbereichen werden in sehr ähnlicher Weise wahrgenommen: Es ist weder für einen Studenten (der etwas lernen und sich nicht verstecken will) noch für einen Unterrichtenden angenehm, in einem Hörsaal mit 500 Zuhörern zu arbeiten. Es ist daher plausibel und vernünftig, dass (zu) große Universitäten in vielen Aspekten (es gibt auch da durchaus Unterschiede) der Lehrsituation schlecht eingestuft werden.

5. Bilden die Antworten auf die ausgewählten Fragen die Lehrleistung ab? Meinefeld (2000: 27) kritisiert, daß ‚... die fachliche Kompetenz des Lehrenden, die Lehrinhalte, die didaktische Aufbereitung des Lehrstoffes oder der Lernerfolg ... gar nicht erfasst ...‘ werden.

Wie schon in Abschnitt 2 dargestellt, ist kein Indikator in der Lage, die Hochschul-Lehrleistung zu messen. Vielleicht wäre es ein sehr gutes Maß für die Lehrqualität, einfach eine hinreichend große Zahl von Studenten aus allen Fachbereichen aller deutschen Universitäten zufällig auszuwählen und einem standardisierten Test zu unterziehen, um dann anhand dessen Ergebnis die Lehrleistung zu messen und zu vergleichen. Dieses Verfahren ist aber so nicht durchführbar. Genauso gibt es keine handhabbaren, allgemein akzeptierten Maße für Lehrinhalte und die fachliche und didaktische Kompetenz der Lehrenden. Daher muss auch Meinefeld (2000: 27) zugestehen, dass die Auswahl der Kriterien den Studien ‚... nicht prinzipiell vorzuwerfen‘ ist. Die Auswahl der Fragen ist mit größtmöglichem Sachverstand durchgeführt worden.

6. Sind nicht alle Ergebnisse aus statistischer Sicht höchst fragwürdig, da hier ordinalskalierte Urteile verschiedener Gruppen von Studenten an verschiedenen Universitäten durch Mittelbildung zusammengefasst und verglichen werden?

Dieses ist natürlich problematisch. Den Vergleich der Urteile verschiedener Gruppen von Studenten kann man nur durch entsprechende Annahmen rechtfertigen. Diese Annahmen können verletzt sein. Das Problem der Mittelbildung über ordinalskalierte Urteile kann man umgehen, indem man - im Einklang mit der statistischen Messtheorie - Mediane statt Mittelwerte verwendet. Die daraus sich ergebenden Konsequenzen für die Ranglisten sollen an anderer Stelle erörtert werden.

7. Ist es sinnvoll, einen Vergleich ganzer Universitäten durchzuführen, bei denen die Leistungen der einzelnen Fachbereiche zusammengerechnet werden?

Diese Vorgehensweise aller SPIEGEL-Ranglisten ist in der Tat unsinnig. Die STERN/START/CHE-Studien verzichten daher auch konsequenterweise darauf: ‚Außerdem ist keinem damit geholfen, wenn man weiß, dass eine Uni mit all ihren Fächern auf irgendeine Weise zusammengerechnet, an neunter Stelle steht ... dem Abiturienten, der Physik studieren will, sagt das gar nichts‘ (Müller-Böling (1999: 23)). Mitglieder eines Fachbereichs haben praktisch keine Einflussmöglichkeit auf die Lehrleistung eines anderen Fachbereichs, die Aspekte ‚guter Lehrbedingungen‘ unterscheiden sich stark zwischen den Fachbereichen (Wirtschaftswissenschaftler brauchen keine Laborplätze wie die Chemiker), die Studenten unterscheiden sich zwischen den Fachbereichen (siehe Punkt 3), usw. Daher sollten sich Universitätsvergleiche auf Fachbereiche beschränken, worauf schon z. B.

Alewell (1988: 47) hinweist. Noch in der Diskussion ist aber, wie fein diese Aufteilung in Fachbereiche sein sollte.

8. Ist es sinnvoll, wie bei der Bildung der Fachbereichsnoten in (2) beschrieben, die studentischen Urteile zu verschiedenen Fragen mit gleichen Gewichten zu einer Fachbereichsnote zusammenzufassen?

Die Autoren der STERN/START/CHE-Studien verneinen dies und liefern nur noch eine Menge von Einzelranglisten, die nicht mehr zusammengefasst werden, geben aber ein ‚Gesamturteil der Studierenden‘ an, ‚wo die Studierenden am zufriedensten sind‘. Es besteht aber ein ‚Rankingbedürfnis‘, wie Müller-Böling und Hornbostel (2000: 81) feststellen, das man mit größtmöglicher wissenschaftlicher Gründlichkeit befriedigen sollte. Man kann es wahrscheinlich vertreten, wie in den SPIEGEL-Ranglisten die studentischen Urteile zu einem Gesamturteil zusammenzufassen. Die identische Gewichtung der Einzelurteile ist aber problematisch, weil sie i. a. nicht der Gewichtung durch die Studenten entsprechen wird. Mein Vorschlag wäre daher, diese Gewichtung durch eine Vorabbefragung zu ermitteln.

9. Ist die Anzahl $S = 18$ der pro Fachbereich befragten Studenten groß genug?

Eine wichtige Gegenfrage ist hier: Groß genug wofür? Wenn man eine wirklich gute Zufallsstichprobe hat, so genügt ein Stichprobenumfang von $S = 18$ zur hinreichend genauen Schätzung des Mittelwertes $\bar{X}_{F,u}$. Sollen aber Angaben gemacht werden zur Genauigkeit einer Rangliste dieser Mittelwerte, so werden höhere Stichprobenumfänge erforderlich sein. Dieses gilt umso mehr, wenn diese Mittelwerte sehr eng beieinander liegen. Dieser Aspekt wird im folgenden ausführlich analysiert.

10. Wie genau sind Fachbereichs-Ranglisten?

Dies ist aus meiner Sicht eine ganz entscheidende Frage. Die erste SPIEGEL-Studie von 1990 veröffentlichte eine Rangliste der Fachbereichsnoten $\bar{X}_{F,u}$ ohne Angaben zur Ungenauigkeit dieser Rangliste und ertete dafür zurecht herbe Kritik. Seitdem aber verfallen alle Produzenten von Ranglisten in das andere Extrem und sind hinsichtlich der Genauigkeit ihrer Ranglisten vorsichtiger als nötig. Man kann, wie im nächsten Abschnitt gezeigt, die Genauigkeit von Ranglisten genauer beschreiben als mit dem im vorigen Abschnitt beschriebenen Ranggruppenverfahren, das durch die Einteilung aller Fachbereiche in nur drei Rangklassen in großem Maße im Datenmaterial vorhandene Informationen vernichtet und zudem noch, wie im SPIEGEL (1993: 156) schon beinahe kurios-süffisant vermerkt, sehr sensitiv ist: ‚Ähnlich wie die Tausendstelsekunde im Sport, kann auch im Ranggruppenmodell der Universitäten ein kleiner Bewertungsunterschied darüber entscheiden, ob ein Fachbereich in das Mittelfeld oder in eine der beiden Extremgruppen gehört‘. Hochsensitive Ergebnisse sind aber aus statistischer Sicht zu vermeiden. Hier kritisiert Meinefeld (2000: 27) auch zurecht die Autoren der STERN/START/CHE-Studie, die die Fachbereiche in eine Rangordnung bringen, aber dann Wert darauf legen (Müller-Böling und Hornbostel (2000: 81)), dass Rangplätze nicht zugewiesen wurden. Die dezente farbliche Markierung wird in der Tat von Laien kaum wahrgenommen.

Das Fazit dieses Abschnitts ist also, dass die Kritiker der Universitätsranglisten teilweise die falschen Aspekte der in den Medien veröffentlichten Universitätsranglisten kritisieren. Wie in Punkt 7 argumentiert, sollten auch die SPIEGEL-Ranglisten auf den Vergleich ganzer Universitäten verzichten. Wenn man wie in den SPIEGEL-Ranglisten studentische Urteile zu verschiedenen Fragen zu einer Fachbereichsnote zusammenfasst, sollte man sich - wie in Punkt 8 erörtert - Gedanken zur Gewichtung der Einzelurteile machen. Beide Punkte sind in der STERN/START/CHE-Studie schon berücksichtigt. Weiterhin sind alle veröffentlichten Universitätsranglisten wegen der Mittelbildung über ordinalskalierte Urteile verschiedener Gruppen von Studenten angreifbar (Punkt 6). Dieses Problem wird im folgenden ausgeblendet. Die beiden wichtigen Kritikpunkte, die in dieser Arbeit untersucht werden sollen, sind die erforderliche Anzahl von Studenten pro Fachbereich (Punkt 9) sowie die fehlende Genauigkeit bei der Angabe der Genauigkeit der veröffentlichten Ranglisten (Punkt 10).

Schon seit langem scheint auf jeden Fall in der deutschsprachigen wissenschaftlichen Literatur eine deutliche Abneigung gegen Ranglisten vorzuherrschen, über die z. B. Alewell (1988: 51) schreibt: ‚... dürften eher einem gewissen Sensationsstreben ... dienen.‘ Schon zu der Zeit kursierte auch die Idee, dass großer Ungenauigkeit in Ranglisten nur durch die Bildung von Ranggruppen zu entgehen ist: ‚Es ist in der Bundesrepublik unmöglich, den Universitäten exakte Rangplätze zuzuweisen - die Universitäten können allenfalls in drei Leistungsgruppen eingeteilt werden‘ (Daniel (1988: 102)). Es ist aber weder ein vernünftiges Ziel, ‚exakte‘ Rangplätze zuzuweisen, noch die gerade erst erstellte Rangliste wieder in Ranggruppen verschwinden zu lassen, deren Einteilung an den Gruppenrändern (s. o.) zudem sehr sensitiv ist.

Im nächsten Abschnitt wird daher ein Verfahren zur Bildung von Bootstrap-Rangkonfidenzintervallen vorgestellt, mit der die Genauigkeit von Ranglisten wesentlich exakter zu quantifizieren ist. Außerdem wird in einer Simulation die Leistungsfähigkeit der Bootstrap-Rangkonfidenzintervalle mit der des Ranggruppenverfahrens verglichen.

5. Bootstrap-Rangkonfidenzintervalle

Es ist ein häufiger Fehler, wie etwa Meinefeld (2000: 28) von den Konfidenzintervallen für Fachbereichsnoten direkt auf Konfidenzintervalle für Ranglistenplätze zu schließen: ‚Demnach liegt die ... Unsicherheitsmarge z.B. für die Universität Potsdam, die hier die Note 2.8 erhielt ... bei plus/minus 0.2, woraus sich ein Notenintervall zwischen 2.6 und 3.0 errechnet. Konkret bedeutet dies: In der vorliegenden Rangliste ... kann die Universität Potsdam sowohl den 9. als auch den 28. Rang einnehmen - und es ist statistisch in keiner Weise möglich, eine genauere Verortung innerhalb dieser Spannweite zu begründen.‘ Abgesehen davon, dass in der obigen Aussage die Irrtumswahrscheinlichkeit fehlt, fehlt ein entscheidender zweiter Gedanke beim Übergang von Noten-Konfidenzintervallen auf Konfidenzintervalle für Ranglistenplätze.

Die Länge der Ranglistenplatz-Intervalle hängt von allen Größen ab, die die Länge der Noten-Intervalle steuern (Noten-Streuung, Studentenzahl, eventuelle Verteilungsannahme), aber zusätzlich noch von der Diskriminierbarkeit der Fachbereichsnoten. Je enger die Fachbereichsnoten zusammen liegen, desto ungenauer werden die Ranglisten und desto länger sollten daher die Ranglistenplatz-Konfidenzintervalle werden.

Wie entwickelt man Konfidenzintervalle für Ranglisten? Es gibt eine statistische Teildisziplin, die sich mit der Auswahl und Anordnung von Individuen oder Populationen befasst (vgl. Bechhofer et al. (1995) und die dort erwähnte Literatur). In dieser basieren die getroffenen statistischen Schlüsse aber i. a. auf bestimmten Verteilungsannahmen an die betrachteten Variablen (hier Noten), die hier nicht getroffen wurden und wohl auch nur schwer vernünftig zu treffen wären. Daher muss hier ein nichtparametrischer Ansatz gewählt werden.

In Anlehnung an das in Jensen (2000: 13f) vorgestellte Verfahren soll auch hier ein Bootstrap-Verfahren zur Ermittlung der Ranglisten-Genauigkeit eingesetzt werden. Bootstrap-Verfahren, von Efron (1979) entwickelt, haben die statistische Theorie und Praxis revolutioniert, weil sie auf einer einfachen, eleganten Idee basieren, einfach und breit anzuwenden sind und dennoch hinreichende mathematische Subtilität aufweisen. Eine gute Einführung geben Efron und Tibshirani (1993).

Die einfache und elegante Grundidee der **Bootstrap-Verfahren** kann man wie folgt beschreiben (vgl. Hall (1992: 4)):

1. Beim Ziehen einer Zufallsstichprobe (Sample) aus einer unbeobachtbaren Grundgesamtheit begeht man einen unbeobachtbaren Zufallsfehler.
2. Um diesen analysieren zu können, wiederholt man den Vorgang, indem man aus der Zufallsstichprobe wiederum eine hinreichend große Anzahl von Zufallsstichproben (Resamples) zieht. Da man in dieser zweiten Stufe die Resamples und das Sample, aus dem gezogen wurde, kennt, kann man den Zufallsfehler dort genau quantifizieren. Handelt es sich wirklich um Zufallsstichproben, so ist schließlich der beobachtbare Zufallsfehler in der zweiten Stufe eine gute Approximation für den unbeobachtbaren Zufallsfehler der ersten Stufe.

Man kann so Aussagen über die Genauigkeit von Schätzern machen, ohne Verteilungsannahmen treffen zu müssen, und überlässt dabei - in Zeiten leistungsfähiger Rechner vernünftigerweise - einen Großteil der Arbeit dem PC. Wie läuft die Berechnung konkret ab? Ist B eine hinreichend große Zahl von Resample-Läufen (etwa $B = 1000$ oder $B = 2000$) und sind Resample-Werte durch ‚*b‘ gekennzeichnet, so ermittelt man Bootstrap-Rangkonfidenzintervalle (BRaKI) für die Fachbereichsnoten eines Fachs F wie folgt:

1. Ermittle die $U \cdot S$ studentischen Urteile

$$\bar{X}_{1,F,u}, \dots, \bar{X}_{S,F,u}, \quad u = 1, \dots, U \quad (8)$$

(bei $N > 1$ Fragen nach Aggregation mittels einfacher oder gewichteter Summation - vgl. Punkt 8 des vorigen Abschnitts) sowie die Fachbereichsnoten der Universitäten $\bar{X}_{F,u}$ für alle U beteiligten Universitäten.

2. Für $b = 1, \dots, B$ Resampling-Läufe:

(a) Erzeuge für alle U Universitäten S künstliche studentische Urteile

$$\bar{X}_{1,F,u}^{*b}, \dots, \bar{X}_{S,F,u}^{*b}, \quad u = 1, \dots, U \quad (9)$$

durch S -faches zufälliges Ziehen mit Zurücklegen aus den studentischen Urteilen (8) der jeweiligen Universität. Von den ursprünglichen studentischen Urteilen werden hier i.a. einige mehrfach, einige auch gar nicht auftreten.

(b) Bilde daraus die künstlichen Fachbereichsnoten

$$\bar{X}_{F,u}^{*b} = \frac{1}{S} \sum_{s=1}^S \bar{X}_{s,F,u}^{*b}, \quad u = 1, \dots, U, \quad (10)$$

ordne diese in einer Rangliste an und speichere für jede Universität den jeweiligen Ranglistenplatz

$$r_{F,u}^{*b}, \quad u = 1, \dots, U. \quad (11)$$

3. Man erhält so zeilenweise eine $(B \times U)$ -Ranglistenplatzmatrix R_F^* . Diese Matrix wird nun spaltenweise sortiert, um Rangkonfidenzintervalle für die Universitäten zu erhalten: Für $u = 1, \dots, U$ ist das $(1 - \alpha)$ -Bootstrap-Rangkonfidenzintervall für den Ranglistenplatz $r_{F,u}$ der Universität u das Intervall

$$\left[r_{F,u}^{*\alpha/2}, r_{F,u}^{*(1-\alpha/2)} \right], \quad (12)$$

wobei $r_{F,u}^{*\alpha/2}$ und $r_{F,u}^{*(1-\alpha/2)}$ das $100\alpha/2$ - bzw. $100(1-\alpha/2)$ -Quantil der Bootstrap-Verteilung $r_{F,u}^*$ sind.

Alternativ kann man analog zu Jensen (2000: 205) in der $(B \times U)$ -Ranglistenplatzmatrix R_F^* für jede Universität u die Anzahl der Resampling-Läufe $m(u)$ zählen, in denen diese Universität z. B. auf Platz 1 liegt, um dann damit die Zugehörigkeitsfunktion (membership function)

$$mf : \{[1], [2], \dots, [U]\} \rightarrow [0, 1], \quad [u] \mapsto \frac{m([u])}{B} \quad (13)$$

auf der Menge der gemäß der (ursprünglichen) Rangliste angeordneten Universitäten und die kumulative Zugehörigkeitsfunktion (cumulative membership function)

$$cmf : \{[1], [2], \dots, [U]\} \rightarrow [0, 1], \quad [u] \mapsto \sum_{i=1}^{|u|} mf([i]) \quad (14)$$

zu berechnen. Die Zugehörigkeitsfunktion $mf(u)$ schätzt dann die Wahrscheinlichkeit, mit der Universität u in der Rangliste die beste Universität ist. Mittels der kumulativen Zugehörigkeitsfunktion kann man die kleinste Menge von Universitäten angeben, die mit vorgegebener Wahrscheinlichkeit die beste Universität enthält.

Implementiert in MATLAB (Version 5.2.1) benötigt ein PENTIUM-Rechner für die Erstellung von $U = 32$ Bootstrap-Rangkonfidenzintervallen bei $S = 32$ Studenten und $B = 1000$ Resampling-Läufen weniger als 30 Sekunden. Im nächsten Abschnitt werden die Eigenschaften dieses Bootstrap-Verfahrens und des konkurrierenden Ranggruppenverfahrens untersucht. Die Zugehörigkeitsfunktion wird in der Anwendung des darauffolgenden Abschnitts eingesetzt.

6. Eigenschaften der Verfahren

Welche Anforderungen werden an Konfidenzschätzer gestellt? Ist das eben beschriebene Bootstrap-Verfahren in dieser Hinsicht leistungsfähig? Von welchen Randbedingungen hängt die Erfüllung dieser Anforderungen ab?

Typische Anforderungen an Konfidenzschätzer entnimmt man etwa Rohatgi (1976: 466 ff): Ein Konfidenzschätzer sollte Konfidenzintervalle erzeugen, die u. a. möglichst genau und möglichst kurz sind. Genauigkeit bedeutet dabei, dass das $(1-\alpha)$ -Bootstrap-Rangkonfidenzintervall für den Ranglistenplatz $r_{F,u}$ aus (12) wirklich zu Konfidenzintervallen führt, die im Mittel $(1-\alpha) \cdot 100\%$ der ‚wahren‘ Rangplätze enthalten.

Bei Bootstrap-Verfahren interessiert man sich dabei im Detail für die Konvergenz des Niveaus des Konfidenzschätzers

$$\lim_{S \rightarrow \infty} P(r_{F,u} \in [r_{F,u}^{*\alpha/2}, r_{F,u}^{*(1-\alpha/2)}]) = 1 - \alpha \quad (15)$$

(‚Konsistenz‘: vgl. Shao und Tu (1995: 141)) und für die Konvergenzgeschwindigkeit dieses Niveaus

$$P(r_{F,u} \in [r_{F,u}^{*\alpha/2}, r_{F,u}^{*(1-\alpha/2)}]) = 1 - \alpha + O(S^{-k/2}) \quad (16)$$

(‚Genauigkeit der Ordnung k ‘: vgl. Shao und Tu (1995: 144)) unter Verwendung des Landauschen Symbols $O(\cdot)$ (vgl. Rohatgi (1976: 6)).

Bootstrap-Verfahren sind generell für ihre breite Anwendbarkeit bekannt. Der Nachweis der theoretischen Eigenschaften dieser Verfahren (vgl. u.a. Shao und Tu (1995: 86)) erfordert aber das Vorliegen von Zufallsstichproben und die Tatsache, dass die ermittelten Fachbereichsnoten $X_{F,u}$ konsistente Schätzer für die ‚wahren‘ Fachbereichsnoten $\mu_{F,u}$ sind. Letzteres bedeutet, dass sich durch Befragung einer beliebig großen Anzahl von Studenten ($S \rightarrow \infty$) pro Fachbereich und Uni durch die Mittelbildung in (2) die ‚wahren‘ Fachbereichsnoten $\mu_{F,u}$ ermitteln lassen. Im Punkt 3 von Abschnitt 4 war diskutiert worden, dass man eigentlich nur Aussagen über die Noten der auskunftswilligen Studenten erhält. Man sollte also, vor allem in den CHE-Studien, die Ergebnisse entsprechend interpretieren.

Weiterhin erfordert die Leistungsfähigkeit der Verfahren eine hinreichende Glattheit der verwendeten Statistiken (Shao und Tu (1995: 151)). Die Tatsache, dass hier Aussagen über die Genauigkeit von Rangplätzen (ein diskretes Merkmal) getroffen werden, beeinträchtigt aber nicht die Anwendbarkeit des Bootstrap-Verfahrens an sich, sondern nur die Konvergenzgeschwindigkeit des Konfidenzniveaus (16).

Da theoretische Resultate zur Leistungsfähigkeit des Bootstrap-Verfahrens in diesem nichtparametrischen Problem natürlich nicht zu erhalten sind, soll in einer Simulation die Leistung analysiert werden. Dabei werden die durchschnittliche Genauigkeit (vgl. (15)) und die durchschnittliche Länge der Konfidenzintervalle gemessen. In Tabelle 1 ist das Simulationsdesign skizziert. Es ist nach (15) und (16) zu erwarten, dass die Leistung des Bootstrap-Verfahrens von der Anzahl der Studenten S abhängt. Weiterhin dürften die Standardabweichungen der Fachbereichsnoten $\text{Std}(X)_{F,u}$ aus (3) und die relative Lage der Fachbereichsnoten zueinander (eng zusammen oder weit auseinander) einen Einfluss haben. Schließlich beeinflusst die Anzahl der Resampling-Läufe B den beim Resampling begangenen Zufallsfehler und damit die Genauigkeit des Bootstrap-Verfahrens.

Anzahl Universitäten	U	32
Anzahl Studenten	S	16, 32, 64, 128
„Wahre“ Rangliste der Fachbereichsnoten (Tab. 2)	Liste	1, 2
Standardabweichung der Fachbereichsnoten	Std	U(0.2,0.5), U(0.2,0.8)
Anzahl Fragen	N	1
Studentische Urteile	$\bar{X}_{s,F,u}$	Beta()
Anzahl Resampling-Läufe	B	1000, 2000
Anzahl Simulationsläufe	SIM	1000, 2000

Tabelle 1: Simulationsdesign

Um ein realitätsnahes Design zu erhalten, wurden bei der Simulation zwei Vektoren von 32 Fachbereichsnoten aus konkreten Studien (Daniel und Hornbostel (1993: 170) und STERN/START/CHE (1999: 178)) als Ausgangspunkt gewählt, wobei „Ties“ durch Anfügung von Nachkommastellen gebrochen wurden und der zweiten Rangliste ein Wert hinzugefügt wurde, um ebenfalls 32 Noten zu erhalten.

Platz	Liste 1	Liste 2
1	2.97	1.8
2	3.07	1.81
3	3.1	2.0
4	3.101	2.1
5	3.11	2.2
6	3.13	2.21
7	3.21	2.3
8	3.23	2.31
9	3.24	2.32
10	3.241	2.4
11	3.25	2.41
12	3.251	2.5
13	3.252	2.51
14	3.3	2.52
15	3.32	2.6
16	3.33	2.61
17	3.331	2.62
18	3.34	2.7
19	3.341	2.71
20	3.37	2.72
21	3.38	2.73
22	3.381	2.8
23	3.45	2.81
24	3.51	3.0
25	3.52	3.1
26	3.54	3.11
27	3.56	3.12
28	3.65	3.13
29	3.66	3.2
30	3.69	3.4
31	3.72	3.7
32	3.78	3.8

Tabelle 2: „Wahre“ Ranglisten von Fachbereichsnoten

Tabelle 2 sind die verwendeten ‚wahren‘ Mittelwerte zu entnehmen. Mit einer Spannweite von 0.81 (gegenüber 2.0) liegen bei der ersten Rangliste die Fachbereichsnoten also deutlich näher beieinander. Die Verwendung dieser ‚wahren‘ Fachbereichsnoten als Ausgangspunkt der Simulation hat den Vorteil, dass man untersuchen kann, wie gut die verwendeten Verfahren die ‚wahre‘ Rangliste ‚wiederfinden‘. In der Simulation wurden nun die meisten Kombinationen der in Tabelle 1 angegebenen Werte untersucht. Dabei wurden zunächst die U Standardabweichungen (3) der Fachbereichsnoten zufällig aus einer stetigen Gleichverteilung ermittelt:

$$Std(X)_{F,u} \sim U(0.2,0.5) \quad \text{bzw.} \quad Std(X)_{F,u} \sim U(0.2,0.8) \quad (17)$$

Dann wurden die $S \cdot U$ studentischen Urteile $\bar{X}_{s,F,u}$ (mit $N = 1$ Frage oder schon aggregiert) aus Beta-Verteilungen (auf dem Intervall (1,6)) mit Mittelwerten aus Tabelle 2 und Standardabweichungen aus (17) zufällig ermittelt und anschließend ganzzahlig gerundet. Alle Berechnungen wurden in MATLAB (Version 5.2.1) mit $SIM = 1000$ oder $SIM = 2000$ Simulationsläufen durchgeführt.

Die Abbildungen 1 und 2 präsentieren die Simulationsergebnisse für das Bootstrap-Verfahren. Abbildung 1 zeigt die durchschnittliche Länge der Bootstrap-Rangkonfidenzintervalle als Funktion der Studentenzahl für zwei verschiedene theoretische Konfidenzniveaus und zwei verschiedene Diskriminierbarkeitsgrade (leicht: Fachbereichsnoten-Liste 2 und geringere Standardabweichungen) (schwer: Fachbereichsnoten-Liste 1 und größere Standardabweichungen).

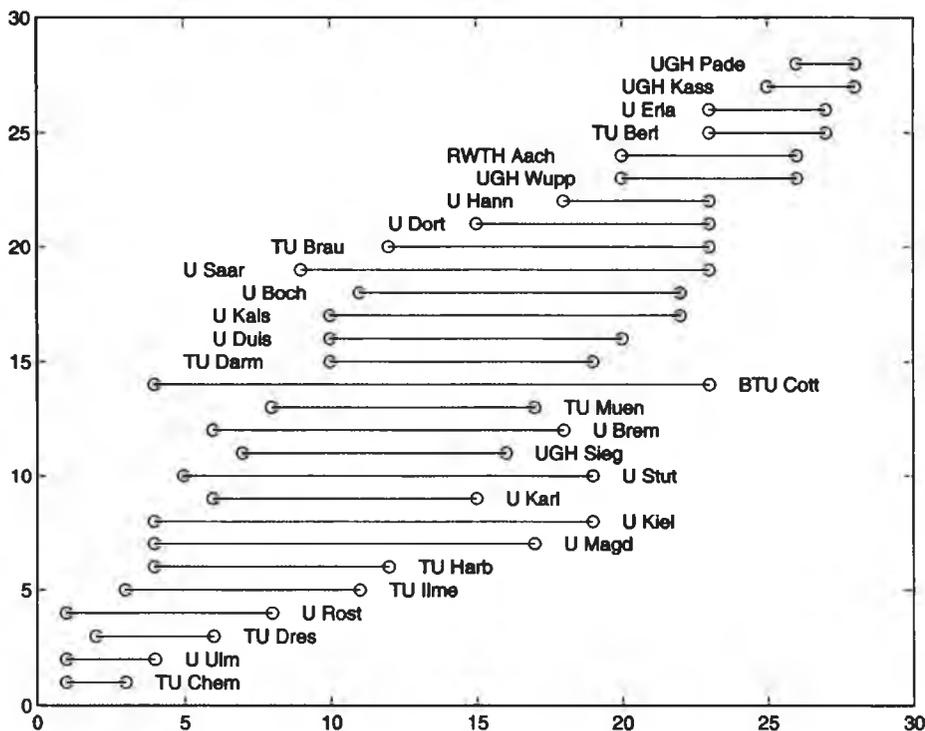
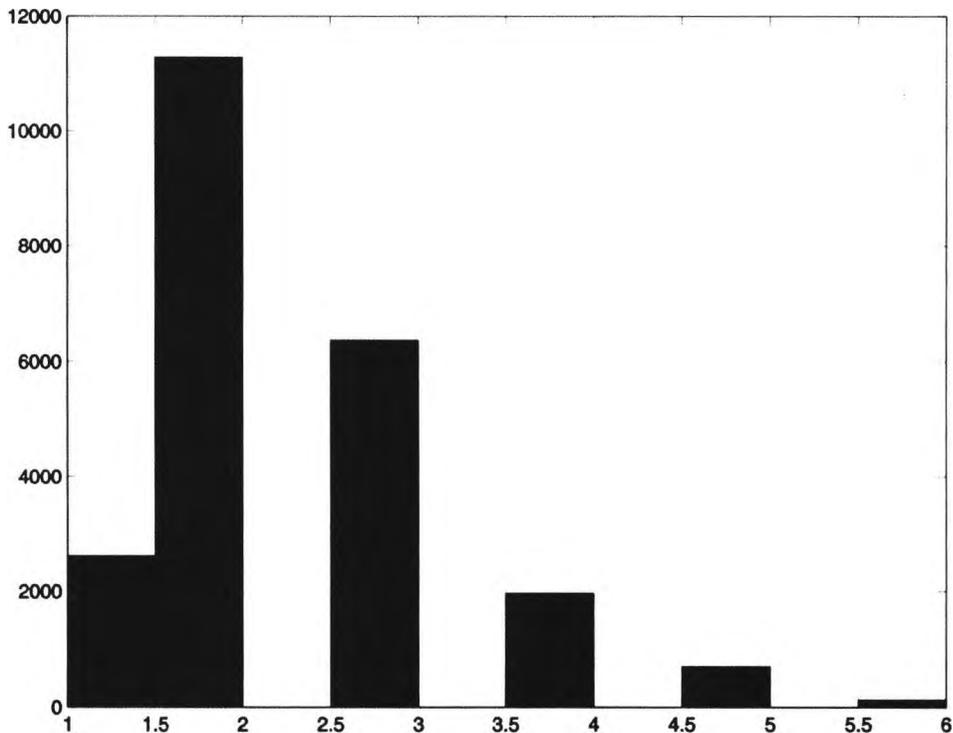


Abbildung 1: Bootstrap-Rangkonfidenzintervall-Länge = $f(\text{Studentenzahl})$

(Niveaus: 0,05 (-), 0,1 (-•); Diskriminierbarkeit: leicht (○), schwer (•))



**Abbildung 2: Bootstrap-Rangkonfidenzintervall-Länge = $f(\text{wahrer Rang})$
(32 Studenten; Niveaus: 0,05 (-), 0,1 (••); Diskriminierbarkeit: leicht)**

Abbildung 1 ist zu entnehmen, dass wachsender Stichprobenumfang S , also steigende Anzahl von Studenten pro Fachbereich plausiblerweise die Länge der Bootstrap-Rangkonfidenzintervalle deutlich senkt (vgl. Abschnitt 4, Punkt 9). Eine Erhöhung der Streuung (Std) der studentischen Urteile und eine schlechtere Diskriminierbarkeit der Fachbereichsnoten (Liste 1 statt 2) bewirken eine deutliche Erhöhung der Intervalllängen.

Abbildung 2 zeigt in typischer Weise die durchschnittliche Länge der Bootstrap-Rangkonfidenzintervalle als Funktion der wahren Ränge in Liste 2 für zwei verschiedene theoretische Konfidenzniveaus, eine Studentenzahl ($S = 32$) und leichte Diskriminierbarkeit (geringere Standardabweichungen). Man sieht, dass in der Mitte der Rangliste, in der die Fachbereichsnoten enger zusammen liegen, die Intervalle wegen der schlechteren Diskriminierbarkeit deutlich länger werden. Aus diesem Grunde sollte man, wenn man sich für die Genauigkeit einer Rangliste interessiert, wie am Beginn des vorigen Abschnitts erläutert, Rang-Konfidenzintervalle und nicht Noten-Konfidenzintervalle betrachten, da letztere diesen Hinweis auf schlechtere Diskriminierbarkeit in der Ranglistenmitte eben nicht enthalten.

leicht diskriminierbar		
α_{theor}	0.05	0.1
α_{BRaKI}	0.041-0.062	0.072-0.096
schwer diskriminierbar		
α_{theor}	0.05	0.1
α_{BRaKI}	0.060-0.082	0.102-0.130

Tabelle 3: Bootstrap-Rangkonfidenzintervalle: Niveaus

Die Erhöhung der Zahl der Resampling-Läufe B von 1000 auf 2000 beeinflusst weder die Genauigkeit noch die Intervalllänge, sodass $B = 1000$ auszureichen scheint. Keiner der verwendeten Simulationsparameter wirkt in klar erkenntlicher Weise auf die Genauigkeit der Bootstrap-Rangkonfidenzintervalle. Die tatsächlichen Konfidenzniveaus α_{BRaKI} weichen in allen betrachteten Konstellationen nicht drastisch, aber doch sichtbar von den theoretischen Niveaus α_{theor} ab (siehe Tabelle 3). Dieses Phänomen dürfte wohl den beiden Diskretisierungen (Übergang von stetigen betaverteilten Zufallszahlen zu ganzzahligen Noten und Übergang von Noten zu Rangplätzen) und u. U. der Verwendung einer festen Rangliste zuzuschreiben sein.

Tabelle 4 zeigt den Anteil der SIM Simulationsläufe, in denen das Ranggruppenverfahren - gemittelt über diese 16 Universitäten - Fachbereiche der (, wahren‘) Plätze 1 bis 8 und 25 bis 32 einer der beiden Extremgruppen zuordnen konnte, was gemäß Daniel (1999: 3) das Ziel des Ranggruppenverfahrens ist.

leicht diskriminierbar		
α_{theor}	0.05	0.1
Zuordnungsanteil	0.22 - 0.29	0.32 - 0.36
schwer diskriminierbar		
α_{theor}	0.05	0.1
Zuordnungsanteil	0.01 - 0.04	0.02 - 0.05

Tabelle 4: Ranggruppenverfahren: Zuordnungsanteil (Ränge 1-8 und 25 - 32)

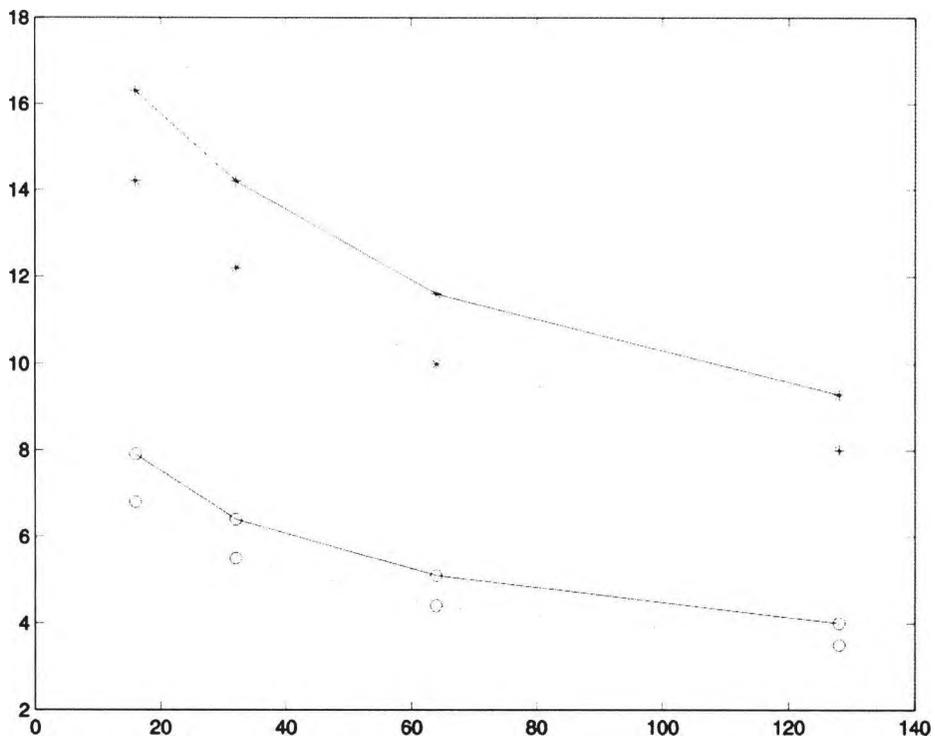


Abbildung 3: Ranggruppenverfahren: Zuordnungsanteil = f(wahrer Rang) 32 Studenten
(32 Studenten; Niveaus: 0,05 (-), 0,1 (+); Diskriminierbarkeit: leicht)

In einigermaßen übersichtlichen Situationen (Liste 2, $Std \leq 0.5$) liefert das Ranggruppenverfahren etwa in 25 % bzw. 33 % der Fälle die Einteilung, die es liefern sollte. In komplizierteren Situationen (Liste 1, $Std \leq 0.8$), in denen das Bootstrap-Verfahren immer noch nützliche Informationen liefert, versagt das Ranggruppenverfahren völlig.

In Abbildung 3 kann man in einer speziellen (günstigen) Konstellation (32 Studenten, leichte Diskriminierbarkeit) den Anteil der SIM Simulationsläufe sehen, in denen das Ranggruppenverfahren die Fachbereiche einer der beiden Extremgruppen zuordnen konnte. Man sieht z. B., dass das Ranggruppenverfahren auch die in Wahrheit beste Universität nur in knapp 70 % aller Fälle erkennt. Das Ranggruppenverfahren liefert also viel weniger Informationen als Bootstrap-Rangkonfidenzintervalle und sollte daher m.E. nicht weiter verwendet werden.

7. Anwendung der Verfahren

Im vorigen Abschnitt sind die Nachteile des Ranggruppenverfahrens gezeigt worden. Daher kommt nun das Bootstrapverfahren in zwei Problemen zum Einsatz.

7.1 Anwendung auf künstliche Daten

Die erste Anwendung dient dazu, Frage 9 aus Abschnitt 4 (Ist die Anzahl der pro Fachbereich befragten Studenten groß genug?) zu beantworten. Daher wird in diesem Falle

mit simulierten Daten gearbeitet, da man so die Anzahl der Studenten pro Fachbereich selbst steuern kann.

Dazu werden die schon im letzten Abschnitt verwendeten $U = 32$ Fachbereichsnoten (Liste 1) aus der SPIEGEL-Rangliste für Betriebswirtschaftslehre von 1993, zu der in Daniel und Hornbostel (1993: 170) nähere Informationen zu finden sind, verwendet. Tabelle 5 gibt die zugehörigen Universitäten und die Ranggruppeneinteilung an. Die jeweiligen Standardabweichungen der Fachbereichsnoten wurden auch aus Daniel und Hornbostel entnommen. Die studentischen Urteile $\bar{X}_{s,F,U}$ für $N = 1$ Frage wurden dann jeweils 100-mal für $S = 18; 32; 64; 128$ künstliche Studenten pro Fachbereich - wie im vorigen Abschnitt - aus auf das Intervall (1,6) transformierten und ganzzahlig gerundeten betaverteilten Zufallszahlen (mit den Mittelwerten und Standardabweichungen aus Daniel und Hornbostel) erzeugt. Die jeweiligen Bootstrap-Rangkonfidenzintervalle mit dem Niveau $\alpha = 0.1$ ergaben sich dann durch Mittelung über diese 100 Simulationsläufe.

Ranggruppe	Universität	Note	
Spitzengruppe	U Düsseldorf	2.97	
	U Mainz	3.07	
	U Stuttgart	3.1	
	U Trier	3.1	
	U Regensburg	3.11	
Mittelfeld	U Mannheim	3.13	
	U Saarbrücken	3.21	
	U Kiel	3.23	
	UGH Siegen	3.24	
	U Lüneburg	3.24	
	U München	3.25	
	U Münster	3.25	
	U Passau	3.25	
	U Bamberg	3.3	
	U Bayreuth	3.32	
	FU Berlin	3.33	
	Schlussgruppe	U Dortmund	3.33
		U E.-Nürnberg	3.34
		U Bielefeld	3.34
U Gießen		3.37	
U Würzburg		3.38	
U Tübingen		3.38	
U Göttingen		3.45	
TH Aachen		3.51	
TU Berlin		3.54	
U Osnabrück		3.56	
U Frankfurt		3.52	
U Oldenbourg		3.65	
U Köln		3.66	
U Marburg		3.69	
U Augsburg	3.72		
U Hamburg	3.78		

Tabelle 5: Fachbereichsnoten: Betriebswirtschaftslehre/Unis, SPIEGEL (1993)

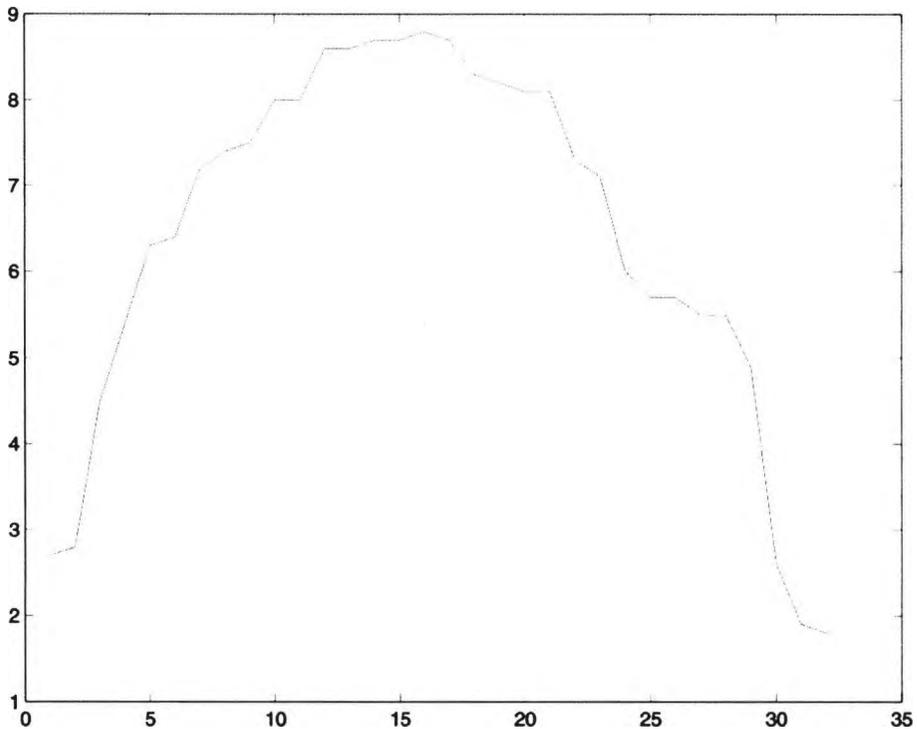


Abbildung 4

Abbildung 4 zeigt das Ergebnis für $S=18$ Studenten, die gleiche Studentenzahl wie im SPIEGEL-Originaldatensatz. Auf der Ordinate sind die Ränge aus der SPIEGEL-Rangliste abgetragen, auf der Abszisse die Ränge für die Bootstrap-Intervalle. Man erkennt, dass - wie im vorigen Abschnitt beschrieben - die Rangkonfidenzintervalle, anders als die Noten-Konfidenzintervalle, in der Ranglistenmitte deutlich länger werden. Man sieht durch das Ziehen einer senkrechten Linie im ‚mittleren Rang‘ 16,5, dass man auch mit dieser Graphik die Ranggruppeneinteilung aus Tabelle 5 reproduzieren könnte.

Aber vor allem sieht man an der weiten Überlappung der Rangkonfidenzintervalle, dass 18 Studenten pro Fachbereich bei weitem nicht ausreichen, um sichere Aussagen über diese Rangliste zu machen. Und gerade an den Rändern der Ranggruppen tut man durch die harte (und hochsensitive) Klassifizierung der Universitäten in eine der Gruppen den Daten (und den Universitäten) Gewalt an. Dass beispielsweise die Universität Frankfurt (FRA) in irgendeiner Weise schlechter eingestuft werden sollte als die Universität Osnabrück (OSN) und die TU Berlin (TBE), leuchtet nach dem Studium der Rangkonfidenzintervalle in keiner Weise ein.

Welche Studentenzahl pro Fachbereich ist dann nötig, um sicherere Aussagen über die Rangliste machen zu können. Bei $S = 32$ Studenten werden die Rangkonfidenzintervalle nur etwas kürzer. Das Ranggruppenverfahren wäre nun in der Lage, 6 Fachbereiche in der Spitzengruppe und 8 in der Schlussgruppe einzuordnen. Bei $S = 64$ Studenten pro Fachbereich wird das Bild noch etwas klarer (vgl. Abbildung 5). Beim Ranggruppenverfahren wären dann 8 Fachbereiche in der Spitzengruppe und 10 in der Schlussgruppe.

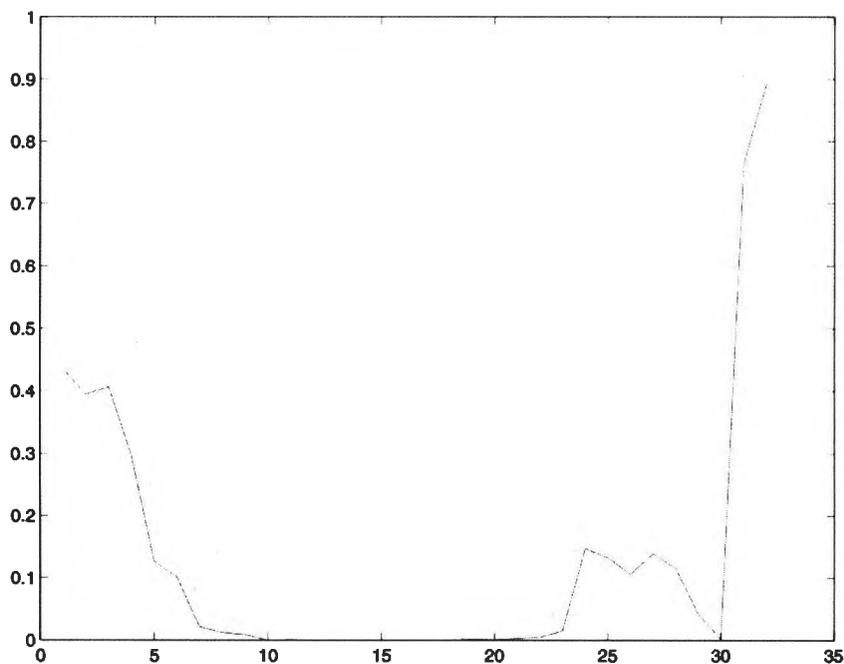


Abbildung 5

Diese Studentenzahl scheint also in dieser Konstellation das mindeste zu sein, um etwas mehr als weit überlappende Rangkonfidenzintervalle, woran die Intervalle wohlgernekt nicht schuld sind, zu erkennen. Die oben erwähnten Vorbehalte gegen das Ranggruppenverfahren bleiben.

7.2 Anwendung auf CHE-Daten

Ranggruppe	Universität	Note
Spitzengruppe	TU Chemnitz	1.5
	U Ulm	1.6
	TU Dresden	1.7
	U Rostock	1.7
	TU Ilmenau	1.8
	TU HH-Harburg	1.9
	U Karlsruhe	2.0
	UGH Siegen	2.0
Mittelfeld	U Magdeburg	1.9
	U Bremen	2.0
	U Kiel	2.0
	TU München	2.0
	U Stuttgart	2.0
	TU Darmstadt	2.1

Ranggruppe	Universität	Note
Mittelfeld	U Duisburg	2.1
	BTU Cottbus	2.1
	U Bochum	2.2
	TU Braunschweig	2.2
	U Kaiserslautern	2.2
	U Saarbrücken	2.2
	U Dortmund	2.3
	Schlussgruppe	U Hannover
RWTH Aachen		2.5
UGH Wuppertal		2.5
TU Berlin		2.6
U Erlangen		2.7
UGH Kassel		2.9
UGH Paderborn		3.0

Tabelle 6: Elektrotechnik/Uni, STERN/START/CHE (2000)

Nun wird das Bootstrap-Verfahren auf den Original-Datensatz für Elektrotechnik an Universitäten aus STERN/START/CHE (2000) angewendet. Tabelle 6 gibt die $U = 28$ Universitäten und für das ‚Gesamturteil der Studierenden‘ (also $N = 1$ Frage) ihre Fachbereichsnoten und Ranggruppenzuordnungen, wie von STERN/ START/ CHE publiziert, an.

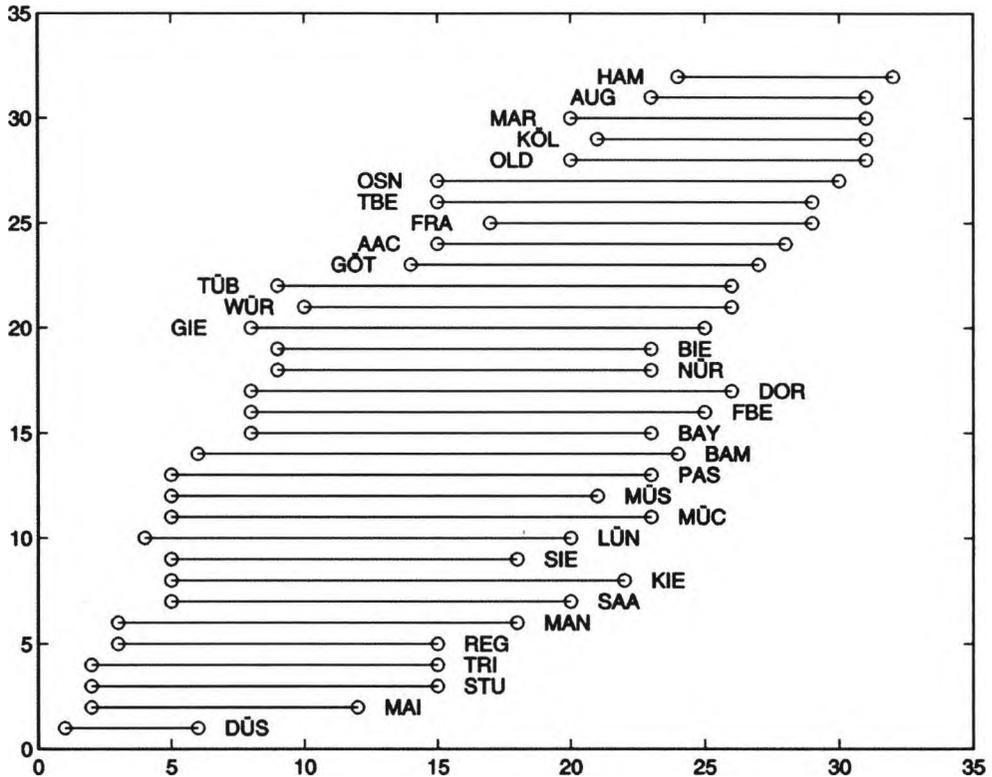


Abbildung 6

Abbildung 6 zeigt die Bootstrap-Rangkonfidenzintervalle für diesen Original-Datensatz. Auf der Ordinate sind wiederum die Ränge aus der CHE-Rangliste abgetragen, auf der Abszisse die Ränge für die Bootstrap-Intervalle. Man erkennt, dass auch hier die Rangkonfidenzintervalle, im Gegensatz zu den Noten-Konfidenzintervallen, in der Ranglistenmitte deutlich länger werden. Zu beachten ist dabei, dass hier die Anzahl der Studenten pro Fachbereich (teilweise stark) schwankt (vgl. die erweiterte Tabelle 7). So liegen von der BTU Cottbus nur $U = 16$ studentische Urteile vor, während z. B. an der Universität Karlsruhe $U = 95$ Studenten geantwortet haben. Daher (und weil die BTU Cottbus in der Ranglistenmitte zu finden ist) ist das Rangkonfidenzintervall der BTU Cottbus außergewöhnlich lang, während das Rangkonfidenzintervall der Universität Karlsruhe, obwohl diese auch in der Ranglistenmitte angesiedelt ist, relativ kurz ist.

Auch in dieser Anwendung fällt die weite Überlappung der Rangkonfidenzintervalle deutlich ins Auge. Und auch hier produziert das Ranggruppenverfahren teilweise sehr unplausible

Ergebnisse, während die Rangkonfidenzintervalle eher in der Lage sind, die Struktur der Rangliste, soweit dieses bei den Studentenzahlen möglich ist, erkennbar zu machen. Warum z. B. die Universität Karlsruhe und die UGH Siegen zur Spitzengruppe gehören sollen, also in irgendeiner Weise mehr mit der TU Chemnitz gemein haben sollen als mit z. B. den Universitäten Magdeburg und Kiel, ist nach dem Studium der Struktur der Rangkonfidenzintervalle höchst unplausibel.

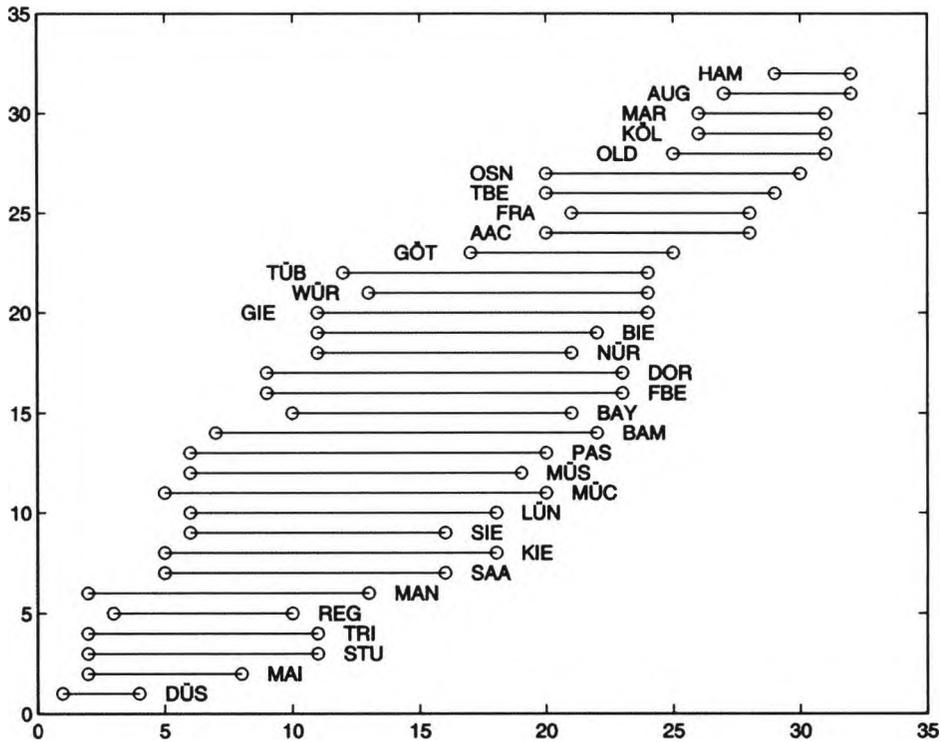


Abbildung 7

Mit Abbildung 7 kann noch ein weiterer Beitrag zur Beantwortung der Frage geleistet werden, warum die vorliegenden Studentenzahlen pro Fachbereich oft nicht für genauere Ranglisten ausreichen. Zu sehen ist nämlich das Histogramm aller 23077 Original-Urteile aller Fächer aller Fachbereiche der STERN/START/CHE (2000)-Studie. An der Tatsache, dass von den 23077 studentischen Urteilen rund 50 % auf die Note 2 und noch einmal mehr als 25% auf die Note 3 entfallen, wird deutlich, dass, wenn nicht sehr viele studentische Urteile pro Fachbereich abgegeben wurden, die Variation der Noten einfach nicht ausreicht, um genauere Ranglisten zu erstellen.

Ranggruppe	Universität	Note	Stud.zahl	Wahrscheinlichkeit für	
				Platz 1	Platz 28
Spitzengruppe	TU Chemnitz	1.5	36	0.606	-
	U Ulm	1.6	73	0.281	-
	TU Dresden	1.7	73	0.040	-
	U Rostock	1.7	25	0.060	-
	TU Ilmenau	1.8	86	0.003	-
	TU HH-Harburg	1.9	82	-	-
	U Karlsruhe	2.0	95	-	-
	UGH Siegen	2.0	75	-	-
Mittelfeld	U Magdeburg	1.9	15	0.005	-
	U Bremen	2.0	40	-	-
	U Kiel	2.0	26	0.003	-
	TU München	2.0	71	-	-
	U Stuttgart	2.0	40	-	-
	TU Darmstadt	2.1	81	-	-
	U Duisburg	2.1	63	-	-
	BTU Cottbus	2.1	16	0.002	0.001
	U Bochum	2.2	55	-	-
	TU Braunschweig	2.2	47	-	-
	U Kaiserslautern	2.2	71	-	-
	U Saarbrücken	2.2	20	-	-
	U Dortmund	2.3	47	-	-
Schlussgruppe	U Hannover	2.4	78	-	-
	RWTH Aachen	2.5	50	-	0.006
	UGH Wuppertal	2.5	41	-	0.004
	TU Berlin	2.6	46	-	0.008
	U Erlangen	2.7	72	-	0.020
	UGH Kassel	2.9	51	-	0.330
	UGH Paderborn	3.0	83	-	0.631

Tabelle 7: Elektrotechnik/Uni, STERN/START/CHE (2000): Erweiterung

Auf der CHE-Tagung in Berlin sind die oben vorgestellten Rangkonfidenzintervalle wegen ihrer Komplexität kritisiert worden. Ich habe zwar m. E. ausführlich genug begründet, warum man aus Abbildung 6 weitaus mehr über die Rangliste lernen kann als aus der sensitiven und stellenweise unplausiblen Ranggruppenzuordnung in Tabelle 6. Trotzdem soll mit den Ergebnissen in Tabelle 7 versucht werden, vielleicht zugänglichere Bootstrap-Resultate zu präsentieren. Dazu wird auf die Zugehörigkeitsfunktion (13) aus Jensen (2000) zurückgegriffen.

Die Nachteile des Ranggruppenverfahrens rühren grundsätzlich daher, dass durch dieses Verfahren überhaupt eine Klasseneinteilung vorgenommen wird. Diese sind häufig informationsvernichtend, sensitiv und an den Klassenrändern unplausibel. Aus diesem Grunde verzichtet man bei statistischen Tests z. B. zunehmend auf die Veröffentlichung der 0/1-Testentscheidung (bei vorgegebenem Testniveau), sondern gibt lieber die zugehörigen p-Werte an. Ein vergleichbares Vorgehen ist mit der Zugehörigkeitsfunktion möglich, die - über das in Abschnitt 5 definierte Bootstrap-Verfahren - die Wahrscheinlichkeit schätzt, mit der eine Universität in der Rangliste z. B. die beste Universität (oder unter den besten 3, etc.) ist.

Tabelle 7 ist Tabelle 6, erweitert um die Studentenzahl pro Fachbereich, die generell zur Beurteilung der Ranglistengenauigkeit angegeben werden sollte, und die mittels der Zugehörigkeitsfunktion geschätzten Wahrscheinlichkeiten, mit der die Universitäten in der Rangliste Platz 1 oder den letzten Platz 28 belegen. Zum besseren Vergleich wurde die (manchmal unplausible) Reihenfolge aus Tabelle 6 beibehalten. Es ist offensichtlich, dass auch in diesen geschätzten Wahrscheinlichkeiten - wie in p-Werten - mehr nützliche Informationen als in der Ranggruppenzuordnung enthalten sind. Und man sieht erneut an den Universitäten Magdeburg und Kiel sowie der schon erwähnten BTU Cottbus, welche Gefahren für die Ranglistengenauigkeit durch zu kleine Studentenzahlen einzelner Fachbereiche entstehen.

8. Schlussfolgerungen

Bei der Messung der Lehrleistung von Hochschulen kann man auf die Evaluation durch Befragung von Studenten nicht verzichten. Schon die Tatsache, dass in den Medien regelmäßig auf diese Weise öffentlichkeitswirksam Hochschul-Ranglisten erstellt werden, macht eine konstruktive wissenschaftliche Analyse dieses Themas dringend erforderlich. Die Vorgehensweise bei der Erstellung der SPIEGEL-Ranglisten ist aus einigen Gründen zu kritisieren. Diese Gründe sind aber nicht diejenigen, die am stärksten in der gegenwärtigen Diskussion stehen.

Alle gegenwärtig veröffentlichten Ranglisten sind angreifbar, da hier ordinalskalierte Urteile verschiedener Gruppen von Studenten an verschiedenen Universitäten durch Mittelbildung zusammengefasst und verglichen werden. Diese Schwäche ist aber nicht einfach zu beheben. Die Autoren der SPIEGEL-Ranglisten sollten auf den Vergleich ganzer Universitäten verzichten und sie sollten sich bei der Zusammenfassung verschiedener studentischer Urteile zu Gesamturteilen Gedanken zur Gewichtung der Einzelurteile machen. Beide Punkte sind in den STERN/START/CHE-Studien schon berücksichtigt.

Entscheidend sind zwei weitere Aspekte. Die Autoren von Ranglisten sollten genauere Angaben zur Genauigkeit dieser Ranglisten machen. Das bisher allgemein angewandte Ranggruppenverfahren vernichtet einen Großteil der in den Daten vorhandenen Information, führt zu hochsensitiven und stellenweise unplausiblen Ergebnissen und ist daher dazu denkbar ungeeignet. Das in dieser Arbeit vorgestellte und näher untersuchte Bootstrap-Verfahren zur Berechnung von Rangkonfidenzintervallen und Zugehörigkeitsfunktionen ist eher in der Lage, die Ungenauigkeit von Ranglisten zu quantifizieren und so den Hunger der Öffentlichkeit nach Ranglisten mit angemessener Genauigkeit zu befriedigen. Schließlich

reicht aber offensichtlich die Anzahl der pro Fachbereich befragten Studenten oft nicht aus, um bei der geringen Variation der erteilten Noten hinreichend genaue Ranglisten erstellen zu können.

Abschließend soll in Übereinstimmung mit Meinefeld (2000: 29) festgehalten werden, dass man die Konsequenzen, die aus einer Rangliste zu ziehen sind, wohl bedenken sollte: ‚... so scheint die Schlussfolgerung nur logisch, durch eine Umverteilung der finanziellen und personellen Mittel diejenigen Universitäten bzw. Fakultäten zu belohnen, die „gute“ Lehre leisten, und diese Mittel bei den „schlechten“ Fakultäten einzusparen. Damit aber erhalten Fakultäten bzw. Universitäten, die im Verhältnis zur Zahl ihrer Studierenden bereits gut ausgestattet sind. . . zusätzliche Mittel, während die Arbeitsbedingungen der anderen weiter verschlechtert werden.‘ Aber die in dieser Arbeit behandelte statistische Aufgabe der Erstellung möglichst genauer Ranglisten ist zunächst einmal unabhängig von der politischen Aufgabe der aus guten Ranglisten zu ziehenden Konsequenzen.

Literatur

- Alewell, K.. „Beurteilung der Leistungen von Hochschulen“. Hrsg. Daniel, H.- D. und R. Fisch. *Evaluation von Forschung*. Universitätsverlag Konstanz. Konstanz 1988. 41 - 58.
- Backes-Gellner, U. „Zum Verhältnis von Forschung und Lehre in sozialwissenschaftlichen Fachbereichen“. Hrsg. Helberger, C. *Ökonomie der Hochschule I*. Duncker & Humblot. Berlin 1989a. 51 - 76.
- Backes-Gellner, U. *Ökonomie der Hochschulforschung*. Gabler. Wiesbaden 1989b.
- Bechhofer, R. E., T. J. Santner und D.M. Goldsman. *Design and analysis of experiments for statistical selection, screening, and multiple comparisons*. Wiley. New York 1995.
- Daniel, H.-D. „Forschungsleistungen wissenschaftlicher Hochschulen im Vergleich“. Hrsg. Daniel, H.-D., und R. Fisch. *Evaluation von Forschung*. Universitätsverlag Konstanz Konstanz 1988. 93 - 104.
- Daniel, H.-D. *Methodenbericht zur SPIEGEL-Hochschul-Studie 1999*. SPIEGEL. Hamburg 1999.
- Daniel, H.-D., und R. Fisch (Hrsg.). *Evaluation von Forschung*. Universitätsverlag Konstanz Konstanz 1988.
- Daniel, H.-D., und S. Hornbostel. „Die Studiensituation in der Betriebswirtschaftslehre im Urteil der Studenten“. *Zeitschrift für Betriebswirtschaft*. Ergänzungsheft 3/93. 1993. 161 - 178.
- Dichev, I.D. „How good are business school rankings?“ *Journal of Business* 72/2. 1999. 201 - 213.
- Efron, B. „Bootstrap methods: another look at the jackknife“, *Annals of Statistics* 7. 1979. 1 - 26.
- Efron, B., und R.J. Tibshirani. *An introduction to the bootstrap*. Chapman & Hall. New York 1993.

- Fisch, R., und H.-D. Daniel. *Messung und Förderung von Forschungsleistung*. Universitätsverlag Konstanz. Konstanz 1986.
- FOCUS. „Wo bitte geht's zur Chefetage?“ *FOCUS* 18/97. 1997. 172 - 181.
- Hall, P. *The bootstrap and Edgeworth expansion*. Springer. New York 1992.
- Helberger, C. (Hrsg.) *Ökonomie der Hochschule*. Duncker & Humblot. Berlin 1989.
- Helberger, C. (Hrsg.) *Ökonomie der Hochschule II*. Duncker & Humblot. Berlin 1991.
- Hornbostel, S. „Der Studienführer des CHE - ein multidimensionales Ranking“. Hrsg. Engel, U., und B. Kirsch. *Qualitätsbeurteilungen von Lehre und Studium*. Campus-Verlag 2000
- Hornbostel, S., und H.-D. Daniel. „Studienbedingungen in der Soziologie“. Hrsg. Artus, H., und M. Herfurt. *Soziologielehre in Deutschland*. Leske und Budrich. Opladen 1996. 11 - 58.
- Jensen, U. *Robuste Frontierfunktionen, methodologische Anmerkungen und Ausbildungsadäquanzmessung* (Habilitationsschrift). Peter Lang. Frankfurt 2001.
- Jensen, U. „Is it efficient to analyse efficiency rankings?“ *Empirical Economics* 25. 2000. 189 - 208.
- Meincke, J. „Ranking für Universitäten?“ *Forschung & Lehre* 7. 1999. 359.
- Meinefeld, W. „Hochschulranking - eine unsichere Basis für Entscheidungen“. *Forschung & Lehre* 1/00. 2000. 26 - 29.
- Müller-Böling, D. „Qualität ist messbar“. *STERN, START und CHE, Die Hochschulcharts*. STERN. Hamburg 1999. 22 - 23.
- Müller-Böling, D., und S. Hornbostel „Fehlinterpretationen und Vorurteile - vom Umgang mit Hochschulrankings und deren Nutzen“. *Forschung & Lehre* 2/00. 2000. 81 - 83.
- Rohatgi, V.K. *An introduction to probability theory and mathematical statistics*. Wiley. New York 1976.
- Schenker-Wicki, A.. *Evaluation von Hochschulleistungen*. Deutscher Universitätsverlag. Wiesbaden 1996.
- Scheuch, E.K. „Wir gut sind unsere Hochschulen?“ *Soziologie* 2. 1990. 73 - 90.
- Shao, J., und D. Tu. *The jackknife and bootstrap*. Springer. New York 1995.
- SPIEGEL. „Welche Uni ist die beste?“ *SPIEGEL SPEZIAL* 1/90. Hamburg 1990
- SPIEGEL. „Welche Uni ist die beste?“ *SPIEGEL SPEZIAL* 3/93. Hamburg 1993
- SPIEGEL. „Deutsche Hochschulen im Vergleich“ *SPIEGEL* 15/99. Hamburg 1999
- STERN, START und CHE. *Die Hochschulcharts*, Hamburg 1999
- STERN, START und CHE. *Der Studienführer*. Hamburg 2000
- Tracy, J. und J. Waldfogel. „The best business schools - a market-based approach“. *Journal of Business* 70/1. 1997. 1 - 31.

Eine Analyse des entscheidungsunterstützenden Informationsgehaltes deutscher und US-amerikanischer Hochschul-Rankings

Christian R. Bayer

Rankings beeinflussen Entscheidungen. Somit ist die Güte eines Ranking in erster Linie an seinem entscheidungsunterstützenden Informationsgehalt zu messen. Dieser wird neben der Wahl „geeigneter“ Indikatoren und statistischer Auswertungsmethoden maßgeblich durch die gewählte Ranking-Methodik bestimmt. In diesem Beitrag wird eine modellgestützte, aus der präskriptiven Entscheidungstheorie abgeleitete, „allgemeine Methodik zur Erstellung von Ranking-Modellen“ auf die vergleichende Bewertung des entscheidungsunterstützenden Informationsgehaltes deutscher und US-amerikanischer Hochschul-Ranking-Studien angewendet. Einige bisher in der Wissenschaft wenig diskutierte methodische Schwächen werden angesprochen und Vorschläge unterbreitet, wie der entscheidungsunterstützende Informationsgehalt von Hochschul-Rankings z. B. durch Einsatz neuer Publikationsmedien wie des Internet gesteigert werden kann.

1 Wozu braucht man ein Hochschul-Ranking?

In der Vergangenheit wurde in der Öffentlichkeit vielfach die Annahme vertreten, bei der Qualität der Lehre existierten keine gravierenden Unterschiede zwischen einzelnen Hochschulen¹ und die Studierberechtigten orientierten sich bei der Wahl ihres Hochschulortes stärker an der Nähe zum elterlichen Wohnort als z. B. an der Konzeption und dem Aufbau des Fachstudienganges oder der Tradition und dem Ruf der Hochschule. Sie wurden in zulassungsbeschränkten Studiengängen durch die Zentralstelle für die Vergabe von Studienplätzen den Hochschulen zugeteilt. Öffentliche Haushaltsmittel wurden durch belastungskennzahlenorientierte Kriterien vergeben, es herrschte in der deutschen Hochschullandschaft eine geringe Wettbewerbsintensität, und ein Leistungsvergleich oder ein Hochschul-Ranking wurden als nicht erforderlich angesehen.

Heute fordert die Politik in Zeiten der Knappheit öffentlicher Mittel eine größere Transparenz hinsichtlich der Qualitäten der Hochschulen in der Lehre und in der Forschung. Eine „neue Akademikergeneration“ beginnt gute Ausbildung nachzufragen, und zwar „ganz gleich wo in Europa“,² Mobilitätsprogramme fördern verstärkt den internationalen Austausch von

¹ „Wurden in der Bundesrepublik noch Unterschiede in der Qualität der Forschung anerkannt, wird die Ausbildungsqualität generell an allen Hochschulen als gleichwertig angesehen. Dies ist nicht zuletzt darauf zurück zu führen, dass in der Bundesrepublik Deutschland die Gleichwertigkeit der Studienabschlüsse (vgl. § 9 Abs. 2 Hochschulrahmengesetz) und damit notwendigerweise der Ausbildung angestrebt wird“ (Hess 1992, S.98).

² Siehe Spiegel (1998a). Das Wissenschaftliche Zentrum für Berufs- und Hochschulforschung der GHS Kassel und die HIS GmbH stellten 1999 durch Umfragen fest: Bei der Studienortwahl überwiegen erstmals fachliche und berufliche Motive gegenüber persönlichen Gründen (Lewin, Heublein, Schreiber und Sommer 1999, WZ 1999, WZ 2000).

Studierenden und Wissenschaftlern,³ und den Hochschulen wird auch in zulassungsbeschränkten Studiengängen vermehrt die Möglichkeit der eigenen Auswahl ihrer Studienanfänger eingeräumt. Die staatlichen Mittelzuweisungen an Hochschulen orientieren sich künftig an der Effizienz und Effektivität der Leistungserbringung in Lehre und Forschung sowie bei der Förderung des wissenschaftlichen Nachwuchses (Bundesministerium für Bildung und Forschung 1998, §5). Die Bewertung der Leistungen ist im Hochschulrahmengesetz des Bundes vom 20. August 1998 noch allgemein durch „eine geeignete Form“ der Evaluation vorgesehen (§6). Das Gesetz über die Hochschulen im Freistaat Sachsen vom 11. Juni 1999, §99, legt konkret die Einführung einer leistungs- und ergebnisorientierten Mittelzuweisung für die Hochschulen fest und fordert von den Hochschulen die Vorlage eines Produkt- und Leistungskatalogs, die Einführung einer Kosten-Leistungsrechnung, die Entwicklung eines kennzifferngestützten Berichtssystems und den Abschluss von Zielvereinbarungen sowohl hochschulintern als auch mit dem Staatsministerium für Wissenschaft und Kunst (SMWK 1999, S.66). Diese Paradigmenwechsel und strukturelle Änderungen⁴ wie:

- stärkere autonome Entscheidungsstrukturen in den Hochschulen,
- die Forderung nach Hochschulen mit messbaren, unterschiedlichen Profilen,
- Hochschul-Evaluationen,
- leistungskennzahlenorientierte öffentliche Finanzierung

führen zu einer Intensivierung des Wettbewerbs im deutschen Hochschulsystem und zu einer *wesentlich höheren Entscheidungskomplexität für die „stakeholder“, also z. B. für die Studierberechtigten, die Unternehmen, die Ministerien und die Öffentlichkeit.*

Ein Beispiel für einen sehr wettbewerbsintensiven Markt im Hochschulbereich ist der US-amerikanische Markt für MBA-Studiengänge. In ihm ist von 1960 bis 1989 die Zahl der Abschlüsse von 5000 auf über 70000 und die Zahl der MBA-Schulen auf über 650 angewachsen.⁵ Die persönliche Informationsbeschaffungs- und -verarbeitungskapazität z. B. eines Studierberechtigten oder eines Personalleiters wird hier durch die zu verarbeitende Informationsmenge wie die unterschiedlichen Qualitäten und Quantitäten von Forschung und Lehre, die verschiedenen Studienbedingungen und Hochschulkulturen, die derzeitigen Arbeitsmarktaussichten für Absolventen und die Details der Studien- und Prüfungsordnungen bei weitem überschritten.⁶

³ Somit wird das deutsche Hochschulsystem zunehmend mit sehr wettbewerbsintensiven internationalen Hochschulsystemen und Leistungsstandards konfrontiert wie dem US-amerikanischen Markt für MBA-Ausbildung im Bereich der Wirtschaftswissenschaften (Chen 1996, D'Aveni 1996).

⁴ Siehe Westdeutsche Rektorenkonferenz (1986b), von Trotha (1993), Bayer (1999).

⁵ Siehe Porter und McKibbin (1988, S.93), Miller (1990), Tracy und Waldfoegel (1997, S.1).

⁶ „Demand for systematic information goes with expansion of opportunities and proliferation of universities. Once the small elite who aspired to university education could - or anyway did - rely on the knowledge of teachers, parents, friends and potential employers for guidance through the limited choices available. But now . . . such informal networks have become inappropiate and unreliable. University education now embraces a broad spectrum of skills, disciplines and activities. Potential students have a range of interests, aspirations and previous experience far wider than before“ (Hodges 1993).

Bei dieser hohen Entscheidungskomplexität entsteht der Wunsch nach Instrumenten, die Leistungen transparent und nach internationalen Standards bewertet darstellen. Dabei werden im Rahmen eines verteilten, multi-kriteriellen Entscheidungsprozesses diejenigen Kriterien, zu deren Erhebung und Bewertung dem Entscheidungsträger die notwendigen Kapazitäten fehlen, an einen Dritten übertragen (Bayer 2001). Entscheidungs- und Bewertungshilfen in Form von Studienführern wie Peterson (2000) und Business School Rankings wie von Business Week (seit 1986) (Jackson 1986) und von US News & World Report (seit 1987) (Solorzano 1987) erlangten trotz erheblicher Widerstände aus den Hochschulen in den USA eine große Popularität.⁷ Sie besitzen heute einen bedeutenden Einfluß auf Image und Reputation von Hochschulen und die Hochschulleitungen sehen einen guten Rangplatz bereits als einen existenziellen Wettbewerbsvorteil an (Hess 1992, Tracy und Waldfogel 1997).

In Deutschland wurden die US-amerikanischen Hochschul-Rankings durch Zeitschriften wie Der Spiegel (1989, 93, 98), Manager Magazin (1990, 92, 94, 95, 96, 97, 98, 99), Der Stern (1993, 99), Focus (1997), Stiftung Warentest in Kooperation mit dem CHE (1998) und stern, start und CHE (1999, 2000) bereits früh kopiert. Sie erlangen trotz unbestrittener methodischer Mängel auch hierzulande einen zunehmenden politischen Einfluß, werden von den Hochschulen bereits als Instrument des Hochschul-Marketing⁸ und von den verschiedenen „stakeholder“ als Instrument zur Entscheidungsunterstützung eingesetzt. Studienanfängerbefragungen an verschiedenen deutschen Hochschulen zeigen, dass sich die Mehrheit der Studienanfänger erst in den letzten sechs Monaten vor der Immatrikulation für eine Hochschule entscheidet und dabei der Ruf einer Hochschule und eines Studienganges, neben der Nähe zum heimatlichen Wohnort, ausschlaggebend sind.⁹

2 Welchen Anforderungen muss ein entscheidungsunterstützendes Ranking genügen?

Ein Ranking und eine Evaluation sind Bestandteile eines verteilten, multi-kriteriellen Entscheidungsprozesses, durch den die Qualitäten und Quantitäten von Leistungen (output, outcomes) und deren Produktionsbedingungen (processes) bewertet werden. Bei Hochschulen

⁷ „Leagues tables are an accepted feature of the higher education landscape in the United States. However antagonistic some institutions are towards rankings - and many are - there is little they can do about them. The system, with its 1300 universities and colleges, is too vast, and the competitive pressures too fierce for any group to boycott them“ (Hodges 1993). „... they became a standard benchmark within the academic community . . . “ (Drew und Karpf 1981). „Wenn Rankings erst einmal bestehen, dann ist es unmöglich, sie zu ignorieren“ (Bradburn 1987). Siehe auch Frackmann (1988, S.143-4,6).

⁸ Siehe <<http://www.hochschul-ranking.de>>.

⁹ Siehe u. a. Klein (2000). Auf die Frage „Aus welchen Gründen studieren Sie an Ihrer jetzigen Hochschule?“ antworten 15 Prozent der von HIS an den Universitäten befragten Studienanfänger des Wintersemesters 1998/99, dass der Grund „Hochschule erhielt für mein Studienfach gute Bewertung in einer Hochschulrangliste/Ranking“ wichtig gewesen sei. 28 Prozent gaben als wichtigen Grund für Ihre Hochschulwahl „guter Ruf von Hochschule und Professoren“ an (Lewin, Heublein, Schreiber und Sommer 1999). „Die Ergebnisse einer Wirkungsstudie zur SPIEGEL-Rangliste 1999 zeigen, dass in allen ZVS-Studiengängen ein enger Zusammenhang besteht zwischen der Platzierung einer Universität nach dem Urteil der Studierenden und der Zu- und Abnahme der Bewerberzahlen“ (Daniel 2001).

handelt es sich hierbei in der Mehrzahl um Dienstleistungen,¹⁰ deren Qualitäten keine Eigenschaften der Leistungen sind, sondern „von außen zum Zwecke der Beurteilung herangetragene, mehrdimensionale Konstrukte darstellen“ (Kromrey 2000). Da Qualitätsaussagen hier Werturteile darstellen, die für verschiedene Zielgruppen unterschiedliche Bedeutungen besitzen, ist die Qualität einer Dienstleistung nicht „objektiv“ in einer Weise empirisch abbildbar, wie es das wissenschaftliche Konzept „Messen“ verlangt.

Die Validität, Reliabilität und die intersubjektive Vergleichbarkeit einer Evaluation bzw. eines Ranking muss durch die Festlegung von Vergleichsmaßstäben bzw. Standards gewährleistet werden (Objektivierung durch ein Evaluations- bzw. Ranking-Verfahren).¹¹ Eine Zielgruppe ist hier eine Gruppe von Leistungsabnehmern, die eine Leistung in einem gegebenem Kontext nach „vergleichbaren“ Zielen bzw. Kriterien beurteilt; wie die Anspruchsgruppen der Studierenden mit oder ohne Vorkenntnisse, mit oder ohne Leistungsmotivation, mit passiv-konsumierenden oder aktiv-entdeckendem Lernstil.¹² Ein Ranking wird definiert als ein Index auf der Menge der Handlungsalternativen einer Menge von Entscheidungsträgern, dessen Indexwerte zu einer, die Präferenzrelationen der Entscheidungsträger repräsentierenden, Rangfolge führen. Dies ist eine notwendige Voraussetzung für ein Ranking, das entscheidungsunterstützende Informationen liefern soll.

Wendet man Klassifikationen von Evaluationsmodellen wie die von Stufebeam und Shinkfield (1985) oder die von Cook und Shadish (1986) auf diese Definition an, so ist sie in die Klasse der wahren Hochschul-Rankings oder der Hochschul-Rankings, die sich am Informationsbedarf der „stakeholder“ orientieren, einzuordnen.¹³ Der subjektive Charakter der Entscheidungssituation einer Ranking-Zielgruppe wird hier bewußt auf das Ranking übertragen, das z. B. einen Studierberechtigten oder einen Personalleiter bei seiner „rationalen“ Wahl

¹⁰ Universitäre Lehrangebote haben immateriellen Charakter, stellen Potentialangebote dar, deren Qualität frühestens zum Zeitpunkt der Inanspruchnahme beurteilt werden kann und erfordern im Rahmen der Leistungserstellung bzw. -inanspruchnahme die Integration und teilweise Mitwirkung der Studierenden im Sinne eines externen Faktors. Die Produktion und der Konsum haben gleichzeitig zu erfolgen. Universitäre Lehrangebote erfüllen somit sämtliche konstitutive Merkmale von Dienstleistungen (Hansen, Henning-Thurau und Wochnowski 1997, S.378).

¹¹ Kromrey (2000b) folgert, dass die Evaluation von Hochschulleistungen in der Methodologie der empirischen Sozialforschung streng genommen nicht möglich ist und definiert in diesem Kontext Evaluation als methodisch kontrollierte, verwertungs- und bewertungsorientierte Form des Sammelns und Auswertens von Informationen.

¹² „Von Studienanfängern und Fortgeschrittenen, von gegenwärtig Studierenden und künftigen Absolventen, von Arbeitgebern und fachwissenschaftlicher community werden unterschiedliche, teils sogar gegensätzliche Erfordernisse geltend gemacht . . . Mit der Festlegung aus wessen Perspektive die qualitätsrelevanten Dimensionen und die Qualitätsindikatoren zu deren Beurteilung zu bestimmen sind werden bereits wesentliche Vorentscheidungen getroffen. Das Qualitätskonzept muss präzise definiert und durch Kriterien, Standards und gültige Indikatoren operationalisiert werden“ (Kromrey 2000).

¹³ Stufebeam und Shinkfield (1985) unterscheiden in Pseudo-Evaluationen, Quasi-Evaluationen und wahre Evaluationen. In der Klasse der Pseudo-Evaluationen befinden sich verdeckte Untersuchungen, die Informationen für z. B. politische Dispute generieren sollen oder Evaluationen zu Zwecken der Öffentlichkeitsarbeit. Die Klasse der Quasi-Evaluationen wird in zielorientierte Evaluationen und experimentelle Arbeiten unterteilt. Die Klasse der wahren Evaluationen besteht aus entscheidungsorientierten Evaluationen, client-centered studies oder stake responsive evaluations (MacDonald 1975, Rippey 1973, Guba 1978). Cook und Shadish (1986) unterscheiden nach Evaluationsmodellen, bei denen eine „machbare Lösung“ gesucht wird (Campbell 1969, Cook und Campbell 1979), kausale Erklärungen gefunden werden sollen (Cronbach 1982, Chen und Rossi 1980, Chen und Rossi 1983, Weiss 1977, Weiss 1978) oder die sich am Informationsbedarf der „stakeholder“ orientieren (Wholey 1960, Stake 1978, Guba und Lincoln 1981).

unterstützen soll. Unter „Rationalität“ wird hier eine prozedurale Rationalität im Sinne der präskriptiven Entscheidungstheorie verstanden (Eisenführ und Weber 1994).

Während die normative Entscheidungstheorie auf einem substantiellen Rationalitätsbegriff basiert, liegt der präskriptiven Entscheidungstheorie ein instrumentaler Rationalitätsbegriff zugrunde, bei dem sich die Erfüllung normativer Regeln nur auf den Entscheidungsprozess und nicht auf die zugrundeliegenden Ziele bezieht (von Nitzsch 1992, S.11-3).

Die Entscheidungshilfe eines Hochschul-Ranking besteht in der zielgruppenspezifischen Strukturierung der Entscheidungssituation, der Beschaffung der entscheidungsnotwendigen Informationen, der zielgruppenspezifischen Auswertung und der konsistenten Verdichtung der Informationen. Dieser Anspruch mag sehr hoch erscheinen; wird er jedoch nicht erfüllt, besitzt das Hochschul-Ranking nur geringen oder keinen entscheidungsunterstützenden Informationsgehalt und es wäre den Klassen Pseudo-Hochschul-Rankings der Quasi-Hochschul-Rankings zuzuordnen.

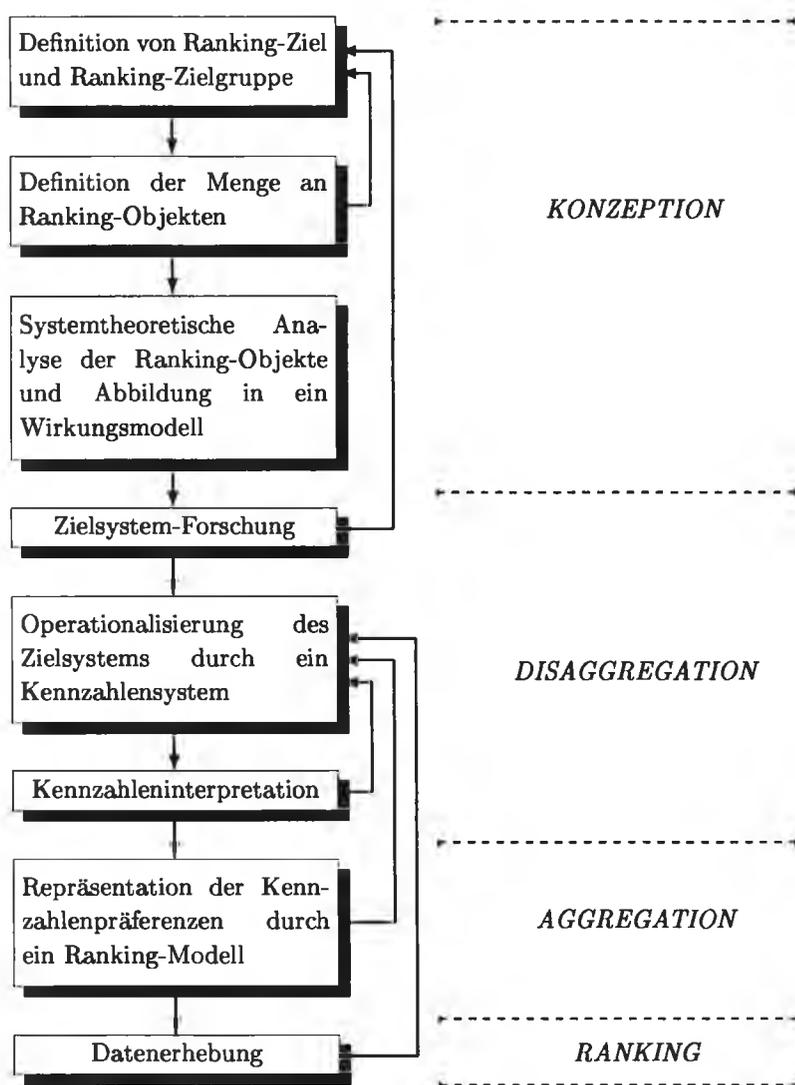


Abb. 1: Allgemeine Methodik zur Erstellung eines Ranking-Modells nach Bayer (1999)

Die in Abbildung 1 dargestellte, „allgemeine Methodik zur Erstellung eines Ranking-Modells“ wurde in Bayer (1999) entwickelt und u. a. durch Kreuzmann (2001) und Syrbe (2001) kommentiert. Sie basiert auf der präskriptiven Entscheidungstheorie und der Evaluationsforschung und bildet die Basis für eine Analyse des entscheidungsunterstützenden Informationsgehaltes deutscher und US-amerikanischer Hochschul-Rankings. Das Ranking-Ziel und die Ranking-Zielgruppe definieren zusammen mit den Ranking-Objekten, d. h. den Wahlmöglichkeiten der Ranking-Zielgruppe bei gegebenem Ranking-Ziel, eine Entscheidungssituation. Auf dem Weg von der Entscheidungssituation zum Ranking kann die Anwendung von Erkenntnissen z. B. aus der Systemtheorie (systemtheoretische Analyse der Ranking-Objekte und Abbildung in ein Wirkungsmodell), aus der präskriptiven Entscheidungstheorie (Strukturierung einer Entscheidungssituation, Zielsystem-Forschung, Repräsentation der Kennzahlenpräferenzen durch Wert- oder Nutzenfunktionen), aus der externen Bilanzanalyse (Anforderungen an Kennzahlen bzw. Kennzahlensystemen), aus der Marketing-Forschung (Datenerhebung) und der angewandten Sozialwissenschaft (Evaluation) sehr hilfreich sein.

3 Analyse des entscheidungsunterstützenden Informationsgehaltes deutscher und US-amerikanischer Hochschul-Rankings

Die untersuchten Hochschul-Rankings unterscheiden sich methodisch in den Ranking-Zielen, den Ranking-Zielgruppen, der Anzahl, der Auswahl und der Gewichtung der Kennzahlen, den Skalen, auf denen die Kennzahlen gemessen werden, der Normierung der Kennzahlen, der Anzahl der Aggregationsstufen und den Aggregationsfunktionen.

3.1 Ranking-Ziel

In den untersuchten deutschen Hochschul-Rankings variieren die Ranking-Ziele mit dem Ranking-Ersteller und dem Erstellungsdatum. Sie umfassen eine Bewertung der Lehrqualität (Spiegel 1989, Spiegel 1993, Spiegel 1998), der Ausbildungsqualität (Manager Magazin 1990), der Forschungsleistungen (Manager Magazin 1990, Fischer und Schwarzer 1992), des Images (Fischer und Schwarzer 1992), der allgemeinen Leistungen (Fischer, Rieker, und Riesch 1994, Rieker und Riesch 1995, Rieker 1996, Sommer 1997, Focus 1997), der Einstellungswahrscheinlichkeit (Gronwald und Wöhrle 1998) oder des Renommées und der Qualität der Ausbildung für wissenschaftliche und praktische Tätigkeit (Stern 1993). Während sich die deutschen Hochschul-Rankings vorwiegend auf die isolierte Bewertung bestimmter Teilbereiche der Qualität akademischer Leistungen beschränken, will US News eine umfassende Bewertung der Qualität akademischer Leistungen amerikanischer Colleges und Universitäten vornehmen.¹⁴ Die Studienführer von Stiftung Warentest in Kooperation mit dem CHE (1998) und stern, start und CHE (1999,2000) wollen den Studienberechtigten eine Orientierungshilfe für die Studienplatzwahl geben (Buhr, Giebisch, Hornbostel und Müller-Böling 2000, Barz, Buhr, Giebisch, Hornbostel und Müller-Böling 1999, Stiftung Warentest 1998).

¹⁴ Siehe US News (1998).

Die Definitionen und die Operationalisierungen der Ranking-Ziele sind von Ranking zu Ranking verschieden und oft nicht explizit. Sie manifestieren sich implizit in den unterschiedlichen Kennzahlensystemen. So verschieden die Ranking-Ziele sind, so unterschiedlich sind die resultierenden Rangfolgen.

3.2 Ranking-Zielgruppe

Hansen, Henning-Thurau und Wochnowski (1997) führten im Frühjahr 1996 eine Erhebung der an wirtschaftswissenschaftlichen Lehrstühlen eingesetzten Verfahren zur Evaluation der Qualität der Lehre durch und kommen zu dem Schluß: „... die existierende Vielzahl an gesellschaftlichen Anspruchsgruppen der Hochschulausbildung“ hat „einen breiten, durchaus nicht durchgängig harmonischen Zielkatalog zur Folge“ (Hansen, Henning-Thurau und Wochnowski 1997, S.377). Sie erheben die Struktur des Konstruktes Qualität der Lehre aus dienstleistungstheoretischer Perspektive durch eine empirische Explorationsstudie und unterscheiden als Anspruchsgruppen (stakeholder) die Anbieter der Dienstleistungen (Professoren im Rahmen ihres Erziehungs- und Bildungsauftrages) und die Nachfrager der Dienstleistungen (Studierende). Die gleiche Differenzierung nehmen auch Hodges (1993), Bathelt und Giese (1995), Rosigkeit (1997) und Hornbostel (2001) vor.¹⁵

Die Frage, inwiefern verschiedene Anspruchsgruppen mit ihren unterschiedlichen Beurteilungsperspektiven und Bewertungskriterien eine Differenzierung eines Hochschul-Ranking nach Ranking-Zielgruppen erfordert, wird in der Literatur zu Rankings kontrovers diskutiert. Die Ranking-Ersteller scheuen die höheren Kosten der Erstellung differenzierterer Hochschul-Rankings und versuchen zu zeigen, dass zwischen verschiedenen Anspruchsgruppen durch anspruchgruppenübergreifende, einheitliche Beurteilungsperspektiven und Bewertungskriterien, keine Interessenkonflikte auftreten oder die Interessenkonflikte „im Aggregat einer repräsentativen Anspruchsgruppe“ vernachlässigbar sind.

Hornbostel (2001) berechnet z. B. den Bravais-Pearsonschen Korrelationskoeffizienten zwischen den Urteilen der Anspruchsgruppen Professoren und Studierende zur Lehrsituation im Studienführer von stern, start und CHE 1999. Bei den Kriterien, die den Studienbedingungen zuzuordnen sind wie Bibliothek, Räume, Computer, Labor, Studienorganisation und Kontakt Studierende - Professoren berechnet er Korrelationskoeffizienten von $r_{XY} \in [0.2, 0.86]$. Bei den Kriterien, die jedoch Qualitätsdimensionen der Lehre betreffen wie die Qualität des Lehrangebots sind die Korrelationskoeffizienten systematisch niedriger ($r_{XY} \in [0.08, 0.39]$). Er verweist weiter auf statistische Untersuchungen wie lineare Regressionsanalysen z. B. zur Validität studentischer Urteile, bei denen kein signifikanter Einfluß von überwiegend demographischen bias-Variablen wie Alter, Geschlecht, Erfahrungsstand und Vorbildung auf die studentische Urteile zur Lehrqualität nachgewiesen werden konnte (Hornbostel 2001). Eine mögliche Ursache für diese Ergebnisse sind methodische Fehler bei der Informationsaufbereitung; die Anwendung von dem Analysezweck nicht angemessenen statistischen Verfahren. So kann z. B. Kromrey (2001) durch eine Sekundäranalyse der Daten

¹⁵ „Angesichts der starken Praxisorientierung der Studierenden in der Rechtswissenschaft . . . ist es leicht nachvollziehbar, dass bei der Bewertung der Lehrsituation die akademischen Erwartungshaltungen der Professoren zu anderen Urteilkriterien führen als die Verwertungsperspektive der Studierenden“ (Hornbostel 2001).

des Studienführers von stern, start und CHE (2000), im Gegensatz zu o. a. Arbeiten, einen signifikanten Einfluss dieser Variablen auf die Studierendenurteile zur Lehrqualität nachweisen. Hornbostel (2001) argumentiert, dass die Berücksichtigung unterschiedlicher Beurteilungsprofile innerhalb einer Anspruchsgruppe wie Studierende unterbleiben kann, wenn die unterschiedlichen Beurteilungsprofile „in allen Fachbereichen mehr oder weniger gleichmäßig auftauchen.“ Das erhobene Aggregat an unterschiedlichen Beurteilungsprofilen weist dann keine systematischen Verzerrungen auf. Dieses Aggregat als allgemein akzeptierten Maßstab für die Qualität der untersuchten Leistung(en) zu interpretieren, ist bedenklich. Eine derartige „Vermischung“ der unterschiedlichen Beurteilungsprofile bzw. Zielsysteme zu einem heterogenen Aggregat ist ein statistisches Artefakt (Kromrey 2001), das hohe Informationsverluste und -verzerrungen für alle Anspruchsgruppen zur Folge haben kann.

Diese Ausführungen zeigen die Notwendigkeit, Hochschul-Rankings, die Qualitäten akademischer Leistungen bewerten, nach Ranking-Zielgruppen, mit homogenen Beurteilungsperspektiven und Bewertungskriterien, zu differenzieren. Statistische Auswertungen von z. B. Studierendenbefragungen zur Qualität der Lehre und Arbeitgeberbefragungen zu ihren Einstellungskriterien bei Hochschulabsolventen zeigen, dass i. d. R. auch innerhalb einzelner Anspruchsgruppen wie Studierende und Unternehmen ein „Perspektivenpluralismus“ existiert und „systematische Unterschiede im Urteilsverhalten verschiedener Studentengruppen . . . zu erheblichen Verzerrungen der zu erfassenden Studiensituation führen“ können (Daniel und Hornbostel 1993). Kromrey (1992), Treinen (1993), Landeck (1994), Bankhofer und Hilbert (1995), Kromrey (2001) kommen durch Cluster-Analysen zu dem gleichen Ergebnis.¹⁶ Bathelt und Giese (1995) folgern aus Spiegel (1993): „Die Inhomogenität der Zuhörerschaft verbietet daher, die Beurteilung der Lehrqualität eines Fachbereichs durch additive Zusammenfassung der Einzelurteile von 18 befragten Studierenden mittels einer Durchschnittsbildung durchzuführen.“ Fürstenberg (1995) fordert: „Weichen die Einzelurteile stark voneinander ab, müssen Beurteilungsprofile gebildet werden“. Kromrey (2001) fordert u. a. dass „Lehrqualität sinnvoll nur relational - als Angemessenheit des Angebots (der Lehrenden) für definierte 'Kunden' (Studierende) - entwickelbar ist.“

Bei einer Evaluation wird durch die Aggregation über sehr unterschiedliche Beurteilungsperspektiven und Bewertungskriterien verschiedener Anspruchsgruppen ein

¹⁶ Eine Vorlesungsbefragung von Studierenden an der Ruhr-Universität Bochum ergab z. B. sehr inhomogene studentische Bewertungen, die durch eine Cluster-Analyse in homogene Bewertungsprofile gegliedert werden konnten (Kromrey 1992, Treinen 1993). Landeck (1994) unterscheidet durch eine Cluster-Analyse vier unterschiedliche Lerntypen und leitet für jeden Lerntyp eine zielgruppengerechte Studienberatung ab. Bankhofer und Hilbert (1995) identifizieren durch eine Cluster-Analyse zehn unterschiedliche Unternehmens-Segmente in Hinblick auf ihre Einstellungskriterien. Eine Aufspaltung der erhobenen Werturteile von Managern in die zwei Gruppen Topmanagement und Berufseinsteiger ergab beim Hochschul-Ranking des Manager Magazins 1994 zwei extrem unterschiedliche Rankings (Fischer, Rieker und Riesch 1994). Leitow (1996) unterscheidet Studierende nach den Studiertypen Minimalisten, Schlenderer, Motivierte, mittlere Studierende, Sprinter, Ambitionierte und Maximalisten. Kromrey (2001) identifiziert durch eine Cluster-Analyse neun unterschiedliche Motivationsprofile von Studierenden. Kromrey (2001) „rekonstruiert“ durch eine Sekundäranalyse der Daten des Studienführers von stern, start und CHE (2000) „empirisch voneinander unterscheidbare Bewertungsmuster“ innerhalb der Anspruchsgruppe der Studierenden. Hornbostel (2001) vergleicht die Korrelationen zwischen den mittleren Studierendenurteilen zur Studiensituation aus Spiegel (1999) und aus CHE (1999) im Fachbereich Mathematik. Seine Hypothese für die Ursache der geringen Korrelation (Bravais-Pearsons $\rho_{XY} = 0.301$, Kendalls $\tau = 0.214$) basiert auf unterschiedlichen Beurteilungsprofilen von Lehramts- und Diplomstudierenden.

möglicherweise hoher Informationsverlust in Kauf genommen. Für ein entscheidungsunterstützendes Hochschul-Ranking darf eine solche Aggregation nicht vorgenommen werden. Ranking-Zielgruppen müssen aus Anspruchsgruppen oder -teilgruppen gebildet werden, deren Beurteilungsprofile innerhalb einer Gruppe einen hohen Homogenitätsgrad und zwischen verschiedenen Gruppen einen geringen Homogenitätsgrad besitzen.

In der Literatur zur Marketingforschung besteht z. B. Konsens über die Notwendigkeit, die unterschiedlichen Beurteilungsperspektiven und Bewertungskriterien verschiedener Anspruchsgruppen durch Kunden- und Marktsegmente zu repräsentieren. Im Markt für PKW-Ersatzreifen wird z. B. zwischen dem sicherheitsbewußten, dem preisbewußten, dem sportlichen und dem ökonomischen Reifenkäufer unterschieden (Bayer 1995). Jeder dieser verschiedenen Käufertypen beurteilt die Qualität eines Reifens nach unterschiedlichen Kriterien. Der Versuch, die Präferenzen sowohl des sportlichen als auch des ökonomischen Reifenkäufers durch einen Kriterienkatalog zu repräsentieren, liefert weder für den sportlichen, noch für den ökonomischen Reifenkäufer eine (kauf-) entscheidungsunterstützende Information, selbst wenn die sportlichen und ökonomischen Reifenkäufer in jeder Stadt denselben prozentualen Anteil besitzen.

Sowohl bei den deutschen Hochschul-Rankings als auch bei den US News Rankings werden als Ranking-Zielgruppe u. a. die Studierenden bzw. die Studierberechtigten angesprochen. Der Kreis der Ranking-Adressaten ist damit zu groß. So ist kritisch anzumerken, dass die untersuchten Hochschul-Rankings nur für diejenige (Anspruchsteil-) Gruppe der Studierberechtigten entscheidungsunterstützende Informationen liefern könnten, die die jeweils verwendeten Kennzahlen ihrer Entscheidung zugrunde legen würden. Einen Fragenkatalog bzw. ein Kennzahlensystem, der bzw. das für jeden Studierenden bzw. jeden Studiertyp repräsentativ ist, wird es schwerlich geben können.

3.3 Ranking-Objektmenen

Eine heterogene Nachfrage nach einer Dienstleistung oder einem Produkt führt i. d. R. zu heterogenen Angeboten. Die unterschiedlichen Angebote können ziel- und zielgruppenspezifische, konkurrierende „Modelle“ hinsichtlich des Ranking-Ziels darstellen, die nur aus der Perspektive einer Ranking-Zielgruppe durch ein gemeinsames Kennzahlensystem repräsentiert und mit gleichen Sollwerten bewertet werden können. Hierbei kann ein Ranking der unterschiedlichen Angebote nur zwischen „strukturähnlichen Leistungsprofilen“ vorgenommen werden.¹⁷ Bei PKW-Vergleichen werden z. B. nur Fahrzeuge gleicher Segmente wie Mini, Kleinwagen, Kompakt, Van, Limousine, Cabrio, Sport oder Luxus bewertet, die in ihrer technischen Leistung, in ihrer Ausstattung und ihrem Preis „vergleichbar“ sind.

Ranking-Objektmenen können durch Leistungsprofile gebildet werden, die innerhalb einer Objektmenge einen hohen Homogenitätsgrad und zwischen verschiedenen Objektmenen einen geringen Homogenitätsgrad besitzen (Hess 1992, S.101-2). Sie sind i. d. R. nicht identisch mit

¹⁷ Turner (1987, S.46) bemerkt im Zusammenhang mit der Social Group Theory: „categorization and comparison depend upon each other and neither can exist without the other: the division of stimuli into classes depends upon perceived similarities and differences (comparative relations), but stimuli can only be compared in so far as they have already been categorized as identical, alike, or equivalent at some higher level of abstraction.“

der Menge aller nationaler Hochschulen, da sich die Hochschulen durch verschiedene Fächerspektren unterscheiden. Selbst innerhalb eines Faches oder eines Studienganges sind signifikante Strukturunterschiede zu erwarten, die u. a. aus dem Selbstverständnis und der Positionierung der Hochschule stammen.

Tracy und Waldfogel (1997) bilden z. B. die Ranking-Objektmenge nach Quartilen der Qualität von Studierenden und erhalten vier zielgruppenspezifische Rankings der TOP 20 MBA-Schulen. Elsbach und Kramer (1996, S.453) untersuchen u. a. die „Dimensions of Business School Identities“ US-amerikanischer MBA-Schulen, die unterschiedliche Zielgruppen von Studierenden adressieren. Internationale Positionierungsanalysen wirtschaftswissenschaftlicher Fakultäten ergaben z. B., „dass Spitzenplätze auf der Leistungsachse sowohl von praxisorientierten . . . als auch von theorieorientierten Hochschulen . . . erreicht werden können und somit eine eindimensionale Betrachtung u. U. nicht angemessen ist“ (Simon 1985). Eine Analyse der Pflichtfächer von 25 Business Schools durch die Co-plot Methode ergab eine Klassifizierung der MBA-Programme in sechs strategische Gruppen. Eine Zuordnung der Rangziffern nach Business Week zu den einzelnen MBA-Schulen ergab, dass die TOP 5 MBA-Schulen verschiedenen strategischen Gruppen (Ranking-Objektmenge) angehören und es offensichtlich keine „beste“ Struktur für eine MBA-Ausbildung gibt. Hier wird deutlich, warum die Existenz von verschiedenen Ranking-Objektmenge eine relative Leistungsbewertung zwischen den Hochschulen einer Ranking-Objektmenge erfordert: Sonst ist es theoretisch möglich, dass ein Rangplatz von z. B. 15 den Spitzenplatz in einer Ranking-Objektmenge bezeichnet.

Neben der Anwendung statistischer Verfahren zur Identifikation von Ranking-Objektmenge sind alternative Vorgehensweisen denkbar. Ranking-Objektmenge können bei Hochschulen auch die unterschiedlichen Rahmenbedingungen reflektieren, unter denen die Hochschulen ihre Leistungen erstellen (Westdeutsche Rektorenkonferenz 1986a). Sie können eine „school's membership in selective organizational categories“ kennzeichnen (Elsbach und Kramer 1996, S.442). Ranking-Objektmenge können auch gebildet werden aus Hochschulen, die untereinander um Studienanfänger, Wissenschaftler und finanzielle Mittel konkurrieren (competitor group) oder eine vergleichbare „role and mission“ besitzen (peer group) oder die sich gegenseitig als vergleichbar akkreditieren (aspiration group) oder geographische, juristische oder traditionelle Gemeinsamkeiten besitzen (predetermined group) (Brinkman und Teeter 1987).

Ranking-Objektmenge und die Zugehörigkeit einer Hochschule zu einer bestimmten Ranking-Objektmenge sind zeitlich nicht konstant, sondern abhängig von der sich ändernden Nachfrage und von den strategischen Entscheidungen der Hochschulleitungen (Segev, Raveh, und Farjoun 1999). Doch auch hier gilt: die Gruppenzugehörigkeit zum Entscheidungszeitpunkt ist für den Studierberechtigten entscheidungsrelevant.

In allen untersuchten deutschen Hochschul-Rankings stellen die einzelnen Hochschulfachbereiche die Ranking-Objekte dar. In Focus (1997), Spiegel (1993) und Spiegel (1989) werden die Fachbereichs-Rankings zu einem Hochschul-Ranking aggregiert. US News erstellt Rankings US-amerikanischer Universitäten und Fachbereiche. Business Week erstellt Rankings internationaler Business Schools die eine AACSB Akkreditierung erhalten haben. US News ermöglicht bei den Universitäts-Rankings eine Aufteilung in Ranking-Objektmenge

nach geographischen Faktoren und nach der Finanzierung. Business Week ermöglicht die geographische Aufteilung der Ranking-Objektmenge nach Nationen.

Werden Hochschulen oder Fachbereiche lediglich geographisch abgegrenzt, dann werden möglicherweise Hochschulen, die sich im Rahmen einer Profilbildung auf differenzierte Gruppen von Studierenden spezialisiert haben (ein Beispiel für unterschiedliche Rahmenbedingungen) unzulässigerweise durch dasselbe, einheitliche Kennzahlensystems bewertet.¹⁸

3.4 Modellbetrachtung

Der Zweck von Wirkungsmodellen in „Input-Output“ Darstellung wie Arvidsson (1986), Fisch (1988), Syrbe und Bayer (1997) und Sinz (1998) ist, durch eine stark vereinfachte Abbildungen alle wesentlichen Eingangs-, Prozess- und Ausgangsgrößen (Eigenschaften), sowie deren Verflechtungen und Rückkopplungen (Relationen) vollständig zu erfassen und so zu strukturieren, dass die zur Leistungsmessung geeigneten Eigenschaften identifiziert werden können. Dabei ist der Vereinfachungs- bzw. Detailierungsgrad abhängig von dem Ranking-Ziel, der Ranking-Zielgruppe und den Ranking-Objekten. Das Zielsystem muss später festlegen, welche der Größen für eine Ranking-Zielgruppe entscheidungsrelevant sind. Die Aufdeckung von Wirkungszusammenhängen und Rückkopplungsschleifen dient der Identifikation von Abhängigkeiten, die sich im Zielsystem niederschlagen.

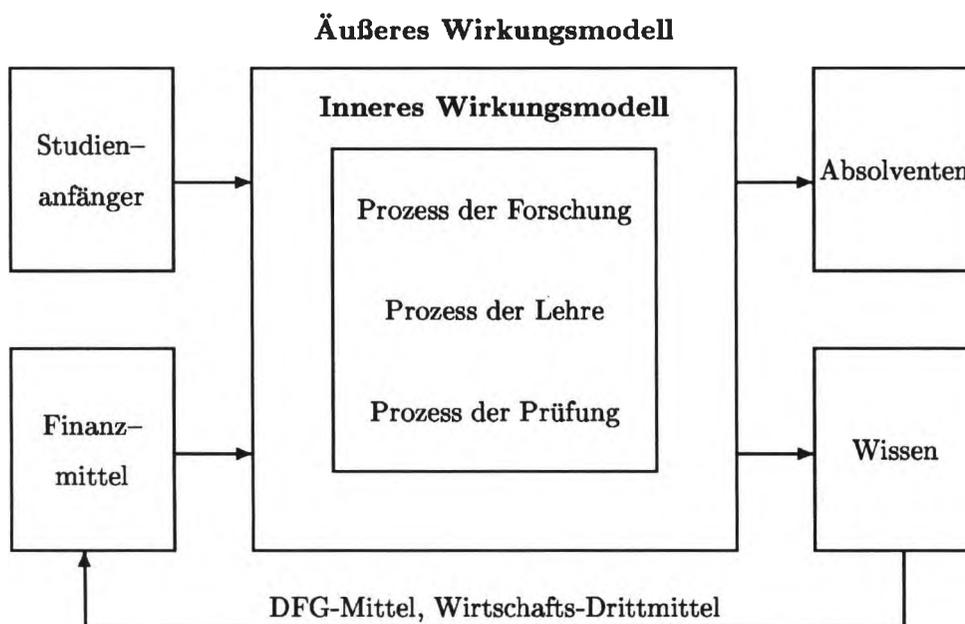


Abb. 2: Quantifizierbares Wirkungsmodell einer Universität nach Syrbe und Bayer (1997)

¹⁸ In Spiegel (1998) wurde lediglich aus Kostengründen eine Vorauswahl der Ranking-Objekte durch eine Professorenbefragung vorgenommen.

Im Unterschied zu den Modellansätzen der 70er Jahre, die durch ressourcenorientierte Kennzahlensysteme die Prozesse einer Universität abbilden und bis ins Detail steuern wollten (z. B. die Berechnung des Personalbedarfs und der maximalen Zulassungszahl),¹⁹ verfolgt das Wirkungsmodell von Syrbe und Bayer (1997) in Abbildung 2 vorrangig ein Strukturierungsziel. Anhand diesem einfachen Strukturmodell wird u. a. klar, dass die Qualitäten der Prozesse Forschung, Lehre und Prüfung nur durch die Qualitäten und Quantitäten der betreffenden Aus- und Eingangsgrößen gemessen werden können und dass die gewöhnlich erhobenen Strukturkennzahlen erst in einem zweiten Schritt, ähnlich einem Benchmarking, das Zustandekommen der unterschiedlichen Prozessqualitäten erklären können.²⁰

Das Wirkungsmodell nach Syrbe und Bayer (1997) und seine Anwendung in der „allgemeinen Methodik zur Erstellung eines Ranking-Modells“ (Abbildung 1) ist in die Theorie der Evaluationsforschung wie folgt einzuordnen: Das Wirkungsmodell modelliert, dem CIPP-Ansatz (Context, Input, Process, Product) nach Stufebeam (1983) folgend, die Universität als Produktionsprozess mit den Eingangsgrößen Studienanfänger und Haushalt & Drittmittel und den Ausgangsgrößen Absolventen und Wissen. Der Output wird durch eine geeignete Kombination der drei Produktionsfaktoren Lehre, Prüfung und Forschung erzeugt. Die Prüfungsleistung wird gesondert aufgeführt, da sie einen großen Einfluß auf die Produktion von Lehre und Forschung hat (Syrbe und Bayer 1997). Im Unterschied zu Stufebeam (1983) wird der Anspruch auf allgemeine Objektivität der Evaluation aufgegeben, da bei den Wertesystemen der Ranking-Zielgruppen eine große Heterogenität zu erwarten ist. Für den Kontext, also die Frage welche Ranking-Zielgruppen müssen mit welchen Bedürfnissen berücksichtigt werden, wird stattdessen wie bei Guba und Lincoln (1989) und Simon (1985) ein konstruktivistischer Ansatz für die Bewertung der Qualität der Eingangs- und Ausgangsgrößen verfolgt.²¹

Hochschul-Rankings, die nicht auf einem Wirkungsmodell basieren, laufen Gefahr, Eingangsgrößen, Prozesse und Ausgangsgrößen unvollständig durch Kriterien bzw. Kennzahlen zu repräsentieren; Prozesse, die in Wirkungszusammenhängen stehen, isoliert zu bewerten; interne Prozesse zu bewerten ohne Berücksichtigung der Eingangs- und Ausgangsgrößen der Hochschule; und Abhängigkeiten zwischen den Kriterien bzw. Kennzahlen nicht zu berücksichtigen. Eine häufige Folge sind auch systematische Fehler in den Bewertungsmodellen, wie die Verwechslung von Struktur- und Qualitätskennzahlen bei der Messung der Qualität des Lehrprozesses.

¹⁹ Das sogenannte FIDES-Kapazitätsmodell (Minke und Weichhold 1972) sollte als Zuordnungsverfahren zur Ermittlung einer „optimalen“ Lehrbelastung dienen. Auf einer ersten Stufe wird die Lehrnachfrage, die durch Studienpläne und die Zahl der Studierenden bestimmt wird, gemäß den Präferenzen des Lehrpersonals, einzelnen Lehrkörpern zugewiesen. Anschließend wurde die Raumauslastung, die sich aus den konkreten, personifizierten Lehrveranstaltungen (und der konkreten Lehrnachfrage durch Studenten) ergibt, berechnet. Ähnliche Modelle werden in Dettweiler und Frey (1972) und Huber (1972) diskutiert.

²⁰ Siehe Frackmann (1988, S.147-8,150-4,7-8), Hamlen und Southwick (1989) und Tracy und Waldfoegel (1997).

²¹ „Jede Gruppe konstruiert aus ihrer Interessenlage und ihrer Sicht der Dinge heraus eine Wirklichkeit. Auch den Evaluatoren gelingt es nicht, eine objektive Rolle einzunehmen. Demzufolge kann ihre Aufgabe nur darin bestehen, die Wirklichkeitskonstrukte der verschiedenen Interessengruppen zutage zu fördern, . . . „ (Kieser, Frese, Müller-Böling, und Thom 1996).

3.5 Zielsystem der Ranking-Zielgruppe

Das Zielsystem einer Ranking-Zielgruppe besteht aus der Menge ihrer möglichst überschneidungsfreien, dekomponierbaren und operationalisierbaren entscheidungsrelevanten Kriterien in einer speziellen Entscheidungssituation, die aus einem Wirkungsmodell der Ranking-Objekte ableitbar ist und in ihrer Gesamtheit die relevanten Eigenschaften der Ranking-Objekte bzgl. des Ranking-Ziels der Ranking-Zielgruppe vollständig beschreibt.

Verfahren zur Ermittlung von Zielsystemen in der multi-attributiven Nutzentheorie unterscheiden zwischen nicht-hierarchischen, hierarchischen und hybriden Verfahren (von Nitzsch 1992, S.69-82). Bei nicht-hierarchischen Verfahren wird eine Liste möglicher Ziele vorgegeben oder empirisch erhoben. Die Ziele müssen auf Redundanz überprüft und entsprechend ihrer Priorität gewichtet werden. Hierarchische Verfahren ermitteln top-down oder bottom-up stufenweise eine Zielhierarchie, indem auf jeder Hierarchieebene Ziele in Unterziele aufgespalten bzw. Ziele zu Oberzielen zusammengefasst werden. Die verschiedenen Verfahren führen in der Regel zu systematisch unterschiedlichen Zielsystemen und sollten kombiniert angewandt werden (von Nitzsch 1992, S.75-6). Stehen mehrere, konkurrierende Zielsysteme zur Auswahl, so empfiehlt es sich, das Zielsystem zu wählen, das der Forderung nach Präferenzunabhängigkeit am ehesten genügt, da sie die Voraussetzung für eine Kennzahleninterpretation durch Einzelpräferenzfunktionen darstellt.

Das Zielsystem z. B. eines Studierenden muss nicht konstant sein und kann sich während des Studiums ändern. Ein zu Studienbeginn pragmatisch orientierter Student (kurze Studienzeit, Berufswunsch Industrie) kann während seines Studiums Gefallen am wissenschaftlichen Arbeiten finden und anschließend eine Karriere in der Wissenschaft anstreben. Entscheidend ist hier jedoch, dass er zum Zeitpunkt seiner Entscheidung einer Ranking-Zielgruppe angehört. Doch bereits das Erkennen der Zugehörigkeit eines Studierberechtigten zu einer definierten Ranking-Zielgruppe trägt entscheidend zur Strukturierung seines Entscheidungsproblems bei. Ändert sich sein Profil oder Zielsystem während des Studiums, so kann der Studierende eine neue Entscheidung unter Berücksichtigung seiner Zugehörigkeit zu einer neuen Ranking-Zielgruppe treffen.

Weder bei den deutschen Hochschul-Rankings noch bei den US News Rankings sind Hinweise auf die Ableitung eines Zielsystems einer Ranking-Zielgruppe bei einer Ranking-Objektmenge aus einer Modellbetrachtung der wichtigsten Eingangsgrößen, Ausgangsgrößen und internen Prozesse der universitären Leistungserstellung zu finden. Ein Beispiel für ein, allerdings willkürliches, top down Verfahren ist das Indikatorenmodell für die Studienwahl des CHE-Studienführers (Barz, Buhr, Giebisch, Hornbostel und Müller-Böling 1999).

3.6 Operationalisierung des Zielsystems durch Kriterien und ein Kennzahlensystem

Erst in Abhängigkeit eines Ranking-Ziels, einer Zuordnung eines Fachbereichs einer Hochschule zu einer Ranking-Objektmenge und eines Zielsystems einer adressierten Ranking-Zielgruppe können Kennzahlen zur Leistungsbewertung aus einem Wirkungsmodell abgeleitet werden (Hindmarsh, Kingston und Loynes 1999). Die Kennzahlen sollen den Zielerreichungsgrad eines Ranking-Objektes bezüglich eines Ziels des Zielsystems der Ranking-Zielgruppe messen. Sie dienen als Indikatoren für einen in der Regel nicht direkt meßbaren Sachverhalt (Indikandum). Die Güte der Indikatoren-Indikandums-Beziehung ist

maßgeblich dafür, ob ein Kennzahlensystem eine geeignete Operationalisierung eines Zielsystems darstellt.²²

Empirische Überprüfungen von Indikatoren-Indikandums-Beziehungen z. B. durch Kausalmodelle der Qualität der Lehre aus Studierendensicht wie bei TEACH-Q (Hansen, Henning-Thurau und Wochnowski 1997) und lineare Strukturgleichungsmodelle wie bei Hornbostel (2001) sind erforderlich; bilden aber die Ausnahme. Oft wird zur Begründung eines Kennzahlensystems auf allgemein akzeptierte Qualitätsmaßstäbe, die sich im Laufe der Zeit oder durch eine Vielzahl von Expertengesprächen „herauskristallisiert“ haben, verwiesen. Die Annahme, dass eine geringe Anzahl von Studierenden je Hochschullehrer studienzeitverkürzend wirkt, konnte z. B. durch Korrelationsanalysen bisher nicht bestätigt werden (Bathelt und Giese 1995).

Da in den untersuchten Hochschul-Rankings weder nach Ranking-Zielgruppen, noch nach Ranking-Objekten differenziert wurde und eine auf einem Wirkungsmodell basierende Ableitung eines Zielsystems der Ranking-Zielgruppe nicht vorgenommen wurde, sind alle Kennzahlensysteme somit nicht begründbar. Um dennoch eine vergleichende Analyse der untersuchten Hochschul-Rankings zu ermöglichen, wird eine nachträgliche, auf dem Wirkungsmodell einer Universität nach Syrbe und Bayer (1997) (siehe Abbildung 2) basierende, Klassifikation der verwendeten Fragen und Kennzahlen in die sieben Kategorien nach Tabelle 1 vorgenommen. Im Folgenden wird vereinfachend für die verwendeten Fragen der Begriff des Kriteriums und für die Antworten auf die Fragen bzw. für die zu den Fragen erhobenen Kennzahlen der Begriff der Kennzahl verwendet.

Die Kennzahlen zu Eingangs- und Ausgangsgrößen werden nach Qualität und Quantität unterschieden (Finanzmittel nach Betrag und Struktur) und die Kennzahlen der Prozesse nach Prozessqualität, Prozessmethoden/ -strukturen und Prozessbedingungen.²³ Die Kennzahlen der Prozessmethoden/ -strukturen beschreiben Prozessprofile und somit die unterschiedlichen Rahmenbedingungen, unter denen Hochschulen ihre Leistungen erstellen. Sie ermöglichen die Identifikation „strukturgleicher“ bzw. „strukturähnlicher“ Prozesse, deren Kenntnis z. B. eine der Voraussetzungen für die Durchführung von Benchmarking- und Ranking-Studien ist. Diese Kennzahlen können eventuell unterschiedliche Qualitäten der Ausgangsgrößen erklären, jedoch nicht messen (Bathelt und Giese 1995, S.139-41). Beispiele für in den Rankings untersuchte Kennzahlen zu Methoden/ Strukturen des Prüfungsprozesses sind die Teilnahme am European Credit Transfer System (ECTS), Teilprüfungen versus Blockexamina und Doppeldiplome. Beispiele für untersuchte Kennzahlen zu Methoden/ Strukturen des Lehrprozesses sind der Umfang des Lehrangebots, der Praxis- oder Theoriebezug der Vorlesungen sowie die Betonung

²² In Turner und Wiswede (1986) werden Beispiele für subjektive Kriterien und objektive Kennzahlen zur Bewertung der Forschung, der Lehre und des Praxisbezugs gegeben. Einen Überblick und eine kritische Diskussion über alternative Wissenschaftsindikatoren gibt Hornbostel (1997, S.180-320).

²³ Hansen, Henning-Thurau, und Langer (2000) unterscheiden beim Qualitätsmanagement von Hochschulen Faculty-Q zwischen der Potentialdimension, der Prozessdimension und der Ergebnisdimension einer Dienstleistung. Die Potentialqualität wird aus der Bewertung der Qualität und Quantität des wissenschaftlichen Lehrpersonals, der Studierenden und der Infrastruktur, die Prozessqualität u. a. aus der Bewertung der Qualität und Quantität des Lehrangebots, der Studierendenbetreuung und des Prozeß- Controlling, die Ergebnisqualität aus der Bewertung der Qualität und Quantität der Abschlüsse, des Außenimages und des Ergebnis-Controlling gebildet.

von Vorlesung versus problemorientierter Gruppenarbeit.²⁴ Prozessbedingungen können die Qualität von Prozessen beeinflussen, jedoch nicht messen. Z. B. haben die Studienbedingungen je nach Studiengang sehr unterschiedlichen Einfluß auf die Qualität des Lehrprozesses.²⁵

Studienanfänger
<i>Qualität</i>
<i>Quantität</i>
Finanzmittel
<i>Betrag</i>
<i>Struktur</i>
• • •
Prozess Forschung
<i>Qualität der Forschung</i>
<i>Methoden/ Strukturen</i>
<i>Forschungsbedingungen</i>
Prozess Lehre
<i>Qualität der Lehre</i>
<i>Methoden/ Strukturen</i>
<i>Studienbedingungen</i>
Prozess Prüfung
<i>Qualität der Prüfung</i>
<i>Methoden/ Strukturen</i>
<i>Prüfungsbedingungen</i>
• • •
Absolventen
<i>Qualität</i>
<i>Quantität</i>
Wissen
<i>Qualität</i>
<i>Quantität</i>

Tab. 1: Kategorien der verwendeten Kennzahlen

²⁴ Tracy und Waldfogel (1997) versuchen durch Prozessgrößen wie „faculty research intensity“, „relative faculty salary“, „percentage of class taught by faculty with Ph.d.s“, „emphasis on case method in teaching“ und „student access to computers“ Unterschiede in der Qualität der Absolventen zu erklären.

²⁵ Sind keine praktischen Übungen bzw. Labortätigkeiten erforderlich, wie z. B. in der Betriebswirtschaftslehre und Volkswirtschaftslehre, und sind gute Vorlesungsskripte vorhanden, so weisen die Studienbedingungen einen weitaus geringeren Einfluss auf die Lehrqualität auf als z. B. in den Studiengängen Medizin, Maschinenbau oder Elektrotechnik.

Die Zuordnung von Kennzahlen zu den Kategorien laut Tabelle 1 kann abhängig sein von der (den) in der Publikation angegebenen Ranking-Zielgruppe(n) und ist nicht immer eindeutig. Die Kennzahl „durchschnittliche Studiendauer“ kann z. B. in Abhängigkeit von der Ranking-Zielgruppe den Kategorien Studienbedingungen (Zielgruppe Studierberechtigte), Prüfungsbedingungen (Zielgruppe Fakultät) oder Qualität der Absolventen (Zielgruppe Wirtschaft) zugeordnet werden. Die folgende Klassifikation und Analyse beschränkt sich deshalb auf die Ranking-Zielgruppe der Studierberechtigten und berücksichtigt die Erläuterung und Interpretation der Kennzahl in der betreffenden Publikation. Sie ist in dem Sinne robust, dass die Aussagen der folgenden Abschnitte auch für modifizierte Zuordnungen der Kennzahlen zu den Kategorien aus Tabelle 1 gültig bleiben und so das notwendige Maß an Willkür bei der Klassifikation vertretbar scheint.

Die Klassifikation der Kennzahlen in die Kategorien nach Tabelle 1 führt zu dem Schluss, dass die untersuchten Kennzahlen bei deutschen Hochschul-Rankings größtenteils nicht dazu geeignet sind, die Hochschulen bezüglich des vorgegebenen Ranking-Ziels zu bewerten. Die folgenden Prozentzahlen geben exemplarisch für die Hochschul-Rankings Spiegel (1989,1993,1998) an, mit welchem Gewicht eine Kategorie in die Gesamtwertung eingeht. Werden z. B. alle Kennzahlen gleich gewichtet, so errechnet sich das Gewicht einer Kategorie aus dem Quotient der Anzahl der Kennzahlen, die dieser Kategorie zugeordnet werden und der Gesamtzahl der Kennzahlen. Eine Kategorisierung der Kennzahlen wird in den ursprünglichen Veröffentlichungen überwiegend nicht vorgenommen. *Eine ausführliche Darstellung der Kennzahlen-Klassifikationen aller untersuchten Hochschul-Rankings ist in Bayer (1999) zu finden.*

	Spiegel	MM	Stern	Focus	US News
Studienanfänger	–	–	–	–	10-25%
Finanzmittel	–	–	22%	–	0-30%
Prozess Forschung	–	22-34%	–	–	0-33%
Prozess Lehre	89%	29-43%	44%	40-48%	0-15%
Prozess Prüfung	0-6%	–	–	0-4%	–
Absolventen	6%	23-26%	22%	13-25%	15-50%
Wissen	0-6%	10%	11%	35-38%	20-32%

Tab. 2: Übersicht über die Gewichtungen der Kennzahlenkategorien in Printmedien

Sowohl die deutschen als auch die US-amerikanischen Hochschul-Rankings geben vor, die Qualitäten akademischer Leistungen zu messen. Einer der wesentlichen Unterschiede zwischen den deutschen Hochschul-Rankings und den US News Rankings liegt darin, dass *die von US News verwendeten Kennzahlen Qualitäten der Ein- und Ausgangsgrößen erfassen und deutsche Hochschul-Rankings vorrangig Studienbedingungen und Methoden/ Strukturen des*

Lehrprozesses, die keine Qualitäts- oder Leistungskriterien darstellen, bewerten (siehe Tabelle 2). Es ist daher auch wenig verwunderlich, dass sich die Sonderauswertungen der Verfasser deutscher Hochschul-Rankings vorrangig auf die Darstellung der Studiensituation beschränken wie Daniel und Hornbostel (1993) und im Ergebnis neu gegründete Fakultäten und Fakultäten mit einer geringen Studierendenzahl aufgrund ihrer besseren Studienbedingungen auf Rangplätzen vor den renommierten Traditionshochschulen liegen. Eine Bewertung der Studienbedingungen (Prozess Lehre) besitzt bei US News im Vergleich mit den deutschen Hochschul-Rankings einen äußerst geringen Stellenwert.

	CHE 1999	CHE 2000	BW 2000	BW 2000 (ext.)
Studienanfänger	–	3%	50%	50%
Finanzmittel	14,3%	3%	–	–
• • •				
Prozess Forschung	–	6%	–	–
Prozess Lehre	85,7%	79%	50%	39%
Prozess Prüfung	–	–	–	–
• • •				
Absolventen	–	–	–	11%
Wissen	–	9%	–	–

Tab. 3: Übersicht über die Gewichtungen der Kennzahlenkategorien in den Suchfiltern

Für die Studienführer von stern, start und CHE (CHE 1999,2000) ergibt eine Klassifikation der Kennzahlen zu denen „individuelle Hitlisten“ durch Suchfilter erzeugt werden können, dass sich diese ebenfalls überwiegend auf den Lehrprozeß und nicht auf die Ein- und Ausgangsgrößen beziehen, die die Leistungen der Hochschulen direkt umfassen. Sie betreffen, nach abnehmender Häufigkeit sortiert, die Studienbedingungen, die Methoden/ Strukturen des Lehrprozesses, die Qualität des Lehrprozesses. Im Unterschied hierzu liegen die Schwerpunkte bei den Suchfiltern von Business Week (BW 2000), nach abnehmender Häufigkeit sortiert, bei der Qualität der Studienanfänger, den Methoden/ Strukturen des Lehrprozesses, den Studienbedingungen, der Qualität der Absolventen. Ein Professorentipp und ein Studierendentipp besitzen bei den Studienführern von stern, start und CHE (CHE 1999,2000) ohne eine Differenzierung nach Ranking-Zielgruppen nur einen geringen entscheidungsunterstützenden Informationsgehalt, und ohne die Klassifizierung der bewerteten Fachbereiche in vergleichbare Ranking-Objektmenen wird hier z. B. die Qualität theoretischer und anwendungsorientierter Forschung durch dieselben Kennzahlen zu schätzen versucht.

3.7 Kennzahleninterpretation

Jede Kennzahl eines Kennzahlensystems muss bezüglich des Ranking-Ziels der Ranking-Zielgruppe bei den Ranking-Objekten und gegebenem Wirkungsmodell aussagekräftig

und einheitlich interpretierbar sein. Eine Kennzahleninterpretation ist also abhängig von dem Ranking-Ziel, der Ranking-Zielgruppe, der Ranking-Objektmenge und von dem Wirkungsmodell. Diese definieren den Kontext, in dem eine Kennzahl interpretiert werden muss. Hierbei werden Kennzahlenwerte durch Präferenzfunktionen auf normierte Wert- bzw. Nutzenskalen abgebildet und somit erst „vergleichbar“ und in diesem Sinn „aggregierbar“ gemacht.²⁶ Die Präferenzfunktionen können z. B. durch Rating-Verfahren, Indifferenz-Verfahren, Trade-Off-Verfahren oder Verfahren der funktionalen Vorbestimmung ermittelt werden (von Nitzsch 1992, S.82-8).

Werden Kennzahlen nicht interpretiert, gehen sie unverändert in das RankingModell ein. Wird zusätzlich, z. B. durch eine lineare Aggregationsfunktion, die Kardinalität der Kennzahl unterstellt, so ist z. B. ein „Anteil ausländischer Studierender“ von 100% zehnmal so „gut“ wie ein Anteil von 10% zu bewerten.²⁷ Eine Interpretation der Kennzahl gäbe die Gelegenheit, einen bestimmten „Sollwert“ maximal zu bewerten und Abweichungen vom Sollwert entsprechend niedriger. Dieses Problem existiert bei vielen Bewertungsfragen wie der Bewertung der Bilanzbonität im Rahmen der externen Bilanzanalyse (Baetge 1994).²⁸

In allen untersuchten Hochschul-Rankings wurden Kennzahlen lediglich normiert, nicht jedoch explizit interpretiert. Für die Fragen, die durch Umfragen erhoben wurden, kann eine implizite Kennzahleninterpretation unterstellt werden, wenn man annimmt, dass sich die Urteile aus der Interpretation einer Mehrzahl von Einzelinformationen bilden. Diese „internen Kennzahlensysteme,“ aus denen die Einzelinformationen gewonnen werden, können individuell sehr verschieden gewesen sein und sind nicht nachvollziehbar.

Eine spezielle Form der Kennzahlennormierung stellt das Ranggruppenverfahren dar. Bei dem Spiegel-Ranking 1993 und bei den Studienführern stern, start und CHE (1999,2000) werden die Hochschulen, die statistisch signifikant besser bewertet werden als andere Hochschulen, lediglich den Ranggruppen Spitzengruppe und Schlussgruppe zugeordnet. Die restlichen Hochschulen bilden die Mittelgruppe. Das Ranggruppenverfahren findet Anwendung bei Umfragen, wenn die erhobenen Beurteilungen größere Standardabweichungen aufweisen und bei Leistungsvergleichen, wenn die Differenzen zu gering erscheinen für die Erstellung einer Rangfolge. Für Aussagen zur „Genauigkeit“ bzw. zur Güte von Ranggruppen schlägt Jensen (2001) vor, anstelle von Ranggruppen Bootstrap-Rangkonfidenzintervallen zu berechnen. Sowohl das Ranggruppenverfahren, als auch das Bootstrap-Verfahren gehen von der Annahme aus, dass aus einer hinreichend großen Zufallsstichprobe an Studierendenurteilen das wahre Urteil durch Mittelwertbildung geschätzt werden kann. Die Heterogenität in den

²⁶ Unterschiedliche Skalen erschweren die Bestimmung der Gewichte von Präferenzfunktionen, da die Gewichte von den Bandbreiten der jeweiligen Skaleneinheiten der Kennzahlenwerte abhängig sind (Bandbreiteneffekt) und deshalb bei der Ermittlung der Präferenzfunktionen mit berücksichtigt werden müssen (von Nitzsch 1992, S.32-7). Durch eine Normierung auf eine einheitliche Skala wie $[0, 1]$ wird der Bandbreiteneffekt eliminiert.

²⁷ Kieser, Frese, Müller-Böling und Thom (1996) bemerken in einem ähnlichen Zusammenhang: „Die Konzentration auf Aktivitäten der Betreuung und der Prozessadministration hat zur Folge, dass Fachbereiche, die 'sich intensiv um ihre Studenten kümmern,' gut abschneiden. Die Annahme, dass Studenten, die sich und ihr Studium gut selbst organisieren können, die praxistauglicheren Absolventen werden, ist jedoch nicht ganz von der Hand zu weisen. Es gibt vielleicht so etwas wie angemessene Informations- und 'Betüterungs'grade, die man nicht überschreiten sollte.“

²⁸ Eine (objektive) Ableitung einer Kennzahleninterpretation kann u. a. durch ökonomische und mathematische Axiome erfolgen (Bayer und Krtscha 1999).

Studierendenurteilen wird durch die Standardabweichung erfaßt und als Zufallsstreuung um das wahre Urteil interpretiert. Die Ursache für die Heterogenität der erhobenen Beurteilungen kann aber auch in den systematisch unterschiedlichen Beurteilungsprofilen verschiedener Anspruchsteilgruppen liegen (vgl. Abschnitt 3.2).

3.8 Repräsentation der Kennzahlenpräferenzen durch ein Ranking

Nach der Interpretation der Kennzahlenwerte liegt für jedes Ranking-Objekt ein Vektor von Einzelinformationen vor, der noch geeignet verdichtet werden muss um ein Ranking erstellen zu können. Grundsätzlich gibt es unendlich viele Möglichkeiten, dies zu tun, und für den Fall, dass kein Kennzahlenvektor einen anderen strikt dominiert, kann jede beliebige Rangfolge durch eine geeignete Wahl der Aggregationsfunktion erzeugt werden. Der Ausweg aus der Beliebigkeit ist zugleich die zwingende Forderung, die an ein Hochschul-Ranking zu stellen ist, wenn es der Ranking-Zielgruppe entscheidungsunterstützende Informationen liefern soll: Die Aggregationsfunktion muss die Kennzahlenpräferenzen der Ranking-Zielgruppe repräsentieren.

Es gibt eine Vielzahl von Verfahren zur Unterstützung von Mehrfachzielentscheidungen (von Nitzsch 1992, S.16-37); in der Anwendung sind jedoch überwiegend die „einfacheren“ Punktbewertungsverfahren und Nutzwertanalysen verbreitet. Dem gegenüber stehen hochentwickelte theoretische Verfahren wie z. B. die multi-attributive Nutzentheorie, die aufgrund ihrer Komplexität bisher wenig Eingang in die Anwendung gefunden haben. Die Präferenzunabhängigkeit im Zielsystem nimmt hier einen zentralen Stellenwert ein. Sie ist die Voraussetzung für die Zerlegung einer Gesamtpräferenzfunktion in (unabhängige) Einzelpräferenzfunktionen und muss notfalls durch eine Redefinition des Zielsystems erzwungen werden (von Nitzsch 1992, S.78-82). Zwischen einer additiven, multiplikativen oder multilinearen Aggregationsfunktion wird durch weitere Unabhängigkeitsaxiome entschieden. Hier können die in jüngster Zeit entwickelten software-gestützten Verfahren helfen, Anwendungsbarrieren zu Überwinden.

In den untersuchten deutschen und US-amerikanischen Hochschul-Rankings wurde in den Fällen, in denen die Aggregationsfunktion explizit angegeben war, eine lineare Aggregationsfunktion des Typs $f(x_1, \dots, x_n) = \alpha_1 x_1 + \dots + \alpha_n x_n$ verwendet. In den anderen Fällen kann die Anwendung einer linearen Aggregationsfunktion vermutet werden. Teilweise erfolgte die Aggregation mehrstufig. Eine notwendige Voraussetzung dafür, dass eine lineare Aggregationsfunktion die Kennzahlenpräferenzen einer Ranking-Zielgruppe repräsentiert und/oder eine Aggregation mehrstufig vorgenommen werden darf, ist die gegenseitige Präferenzunabhängigkeit. Den Nachweis für diese restriktive Annahme bleiben alle Hochschul-Rankings schuldig.

3.9 Datenerhebung

Nach der Bestimmung eines Kennzahlensystems, einer Kennzahleninterpretation und eines die Kennzahlenpräferenzen einer Ranking-Zielgruppe bei gegebenen Ranking-Objekten und Wirkungsmodell repräsentierenden Ranking sind noch abschließend geeignete Informationsquellen für die Erhebung der Kennzahlenwerte festzulegen. „Regardless of how sophisticated a person is at decision-making, his results will be no better than his information“ (Hammond 1965, S.655).

Die „Vergleichbarkeit“ der Kennzahlenwerte der Ranking-Objekte erfordert, dass sie nach einer einheitlichen Methode berechnet und um systematische Verzerrungen bereinigt werden. Z. B. messen Einstiegsgehälter von Hochschulabsolventen kumulativ u. a. regionale, branchen-, berufs- und studierendenspezifische Qualitäten und sind ohne Standardisierung bzw. Bereinigung durch regionale, branchen- und berufsspezifische Indizes und durch die Qualität der Studienanfänger zum Zeitpunkt des Studienbeginns nicht vergleichbar und nicht zur Messung der Qualität einer Hochschulausbildung geeignet (Tracy und Waldfogel 1997, S.17-8). Welsch und Ehrenheim (1999) bereinigen die Anzahl der Promotionen und der Habilitationen bei einem Produktivitätsvergleich volkswirtschaftlicher Fachbereiche in Deutschland um die Größe und das Alter des Fachbereichs (Anzahl Professuren x Betriebsjahre). Mit Ausnahme des Kriteriums Internationalität (Focus 1997) wurden alle in deutschen Hochschul-Rankings bewerteten Kriterien durch Umfragen erhoben. Diese Art der Datenerhebung wurde in der Literatur bereits scharf kritisiert (Bathelt und Giese 1995). Die Kritikpunkte im Einzelnen sind:

- Willkürliche, nicht nachvollziehbare Urteilsbildung bei den Befragten,²⁹
- Verletzung der Forderung nach Repräsentativität der befragten Personen,³⁰
- Mittelwertbildung über intersubjektiv nicht vergleichbare Skalen,³¹
- Verletzung der Forderung nach Reliabilität durch fehlende Überprüfung der Reproduzierbarkeit der Ergebnisse,³²
- geringe statistische Signifikanz der Rangunterschiede.³³

Im Unterschied dazu werden bei den US News Rankings alle Kriterien außer der akademischen Reputation durch quantitative Kennzahlen, die objektiv erhoben werden können, gemessen. Damit treten die bereits diskutierten Probleme der Datenerhebung durch Umfrageforschung bei den US News Rankings in erheblich geringerem Umfang auf. Die Leistungsvergleiche basieren auf objektiv meßbaren Größen wie der Anzahl der Zulassungen im Verhältnis zur Anzahl der

²⁹ „Der Bielefelder Sozialwissenschaftler und Bildungsforscher Professor Klaus Hurrelmann etwa verweist auf die ‚klaren methodischen Beschränkungen‘ der Befragungen: Sie gründen sich auf Stereotype und vorurteilsbehaftete Einstellungen. Urteile werden auf der Basis subtiler und teilweise willkürlicher Bewertungen abgegeben“ (Himmelrath 1997). Siehe auch Spiegel (1993) und Bathelt und Giese (1995, S.141).

³⁰ Ein Reputationsvergleich erfordert die Beurteilung der einzelnen Leistungsdimensionen durch eine repräsentative Stichprobe der Zielgruppe oder durch Experten. Die Repräsentativität muss in den Fällen angezweifelt werden, bei denen Nicht-Experten zu einer Leistungsdimension bzw. einem Kriterium befragt wurden oder die Anzahl der Befragungen statistisch zu gering war (Bathelt und Giese 1995, S.141-8). Siehe auch Bathelt und Giese (1995, S.142), Fürstenberg (1995) und Rosigkeit (1997, S.38). Siehe Bathelt und Giese (1995, S.143-8) und Simon (1985) für den Einfluß und die Auswirkungen von Stichprobenumfang und Zufallsfehler. Die durchschnittliche Rücklaufquote bei Fragebogenaktionen liegt meist unter 30% (oft unter 10%); die Repräsentativität ist somit schwer zu gewährleisten (Daniel 1997).

³¹ Siehe Bathelt und Giese (1995, S.142). Ein Durchschnittswert aus subjektiven Urteilen besitzt nur dann einen empirischen Gehalt, wenn die erhobenen Urteile eine „zentrale Tendenz“ aufweisen (Kromrey 2001). Die Bildung eines Durchschnittswertes ist außerdem nur dann zulässig, wenn die Beurteilungsskalen aller Befragten intersubjektiv vergleichbar sind, d. h. denselben Nullpunkt und die gleiche Maßeinheit besitzen. Die Varianz in den abgegebenen Urteilen kann ein Indiz für die Verletzung der Forderung der intersubjektiven Vergleichbarkeit der Bewertungsskalen sein.

³² Die Reproduzierbarkeit der Rangfolgen wurde z. B. in Stern (1993) durch Stichprobensplits durchgeführt.

³³ Die Antworten in den Stichproben der befragten Personen weisen im Vergleich zum Mittelwert regelmäßig hohe Standardabweichungen auf, so dass viele Rangunterschiede „zufällig“ sein können.

Bewerber, der Anzahl an Diplomen, der durchschnittlichen Punktzahl der zugelassenen Studierenden im Zulassungstest oder der Anzahl von Vollzeit-Doktoranden im Verhältnis zur Anzahl des wissenschaftlichen Personals und nicht auf Durchschnittswerten aus den subjektiven Einschätzungen einer Vielzahl von Professoren, Studierenden und Personalleitern mit sehr unterschiedlichen, individuellen Informationsständen und Zielsystemen.³⁴ Die deutschen Ranking-Ersteller verteidigen die gegenüber dem Leistungsvergleich kostengünstigere Umfrageforschung und verweisen auf die im Vergleich zu den USA mangelnde Verfügbarkeit von aussagefähigen Statistiken. Bei dieser Argumentation wird übersehen, dass der entscheidungsunterstützende Informationsgehalt zwischem einem quantitativen Indikator und subjektiv erhobenen Urteilen sehr verschieden ist. Ein quantitativer Indikator ermöglicht unterschiedlichen Ranking-Zielgruppen zu unterschiedlichen subjektiven Urteilen zu kommen. Werden aber bereits subjektive Urteile erhoben, müssen für den Fall heterogener Beurteilungsprofile diese ermittelt werden. Handelt es sich bei den Urteilen um Qualitäts- oder Akzeptanzurteile, so sind heterogene Beurteilungsprofile eher die Regel als die Ausnahme (Kromrey 2001).³⁵

4 Zusammenfassung und Ausblick auf den CHE- Studienführer

Die Güte eines Ranking ist in erster Linie an seinem entscheidungsunterstützenden Informationsgehalt zu messen. Dieser kann durch die Anwendung eines Ranking-Verfahrens bzw. der vorgeschlagenen „allgemeinen Methodik zur Erstellung von Ranking-Modellen“ (Abbildung 1) bewertet werden. Ein Hochschul-Ranking, das entscheidungsunterstützende Informationen liefern soll, muss ziel- und zielgruppenabhängig erstellt und darf nur in Abhängigkeit des Ranking-Ziels, der Ranking-Zielgruppe, der Ranking-Objektmenge und eines Wirkungsmodells interpretiert werden. Aus den festgestellten methodischen Mängeln deutscher und US-amerikanischer Hochschul-Rankings ergeben sich die folgenden Vorschläge, den entscheidungsunterstützenden Informationsgehalt künftiger Hochschul-Rankings und Studienführer zu erhöhen:

- Die Ziel- und Zielgruppenabhängigkeit eines Hochschul-Ranking macht zunächst ein besseres Verständnis der Ranking-Zielgruppen wie die der Studierberechtigten und der Personalleiter, insbesondere ihrer unterschiedlichen Präferenzstrukturen erforderlich.
- Hochschulen, die sich durch unterschiedliche Profilbildung auf verschiedene Ranking-Zielgruppen spezialisiert haben, dürfen nur in Abhängigkeit von der relevanten Ranking-Zielgruppe bewertet werden (Ranking-Objektmenge).

³⁴ Aus diesen Gründen fordern u. a. Bathelt und Giese (1995) auch deutsche Hochschul-Rankings auf der Basis quantitativer Kennzahlen zu erheben.

³⁵ Weiter ist zu prüfen, ob die Fragen den methodischen Anforderungen der Umfrageforschung genügen. Die Fragen müssen z. B. in dem Sinne eindeutig formuliert werden, dass jeder Befragte unter jeder Frage dasselbe versteht (Fehler bei der Informationssammlung) und die Befragten müssen in der Lage sein, die Frage kompetent zu beantworten. So führt u. a. Kromrey an, dass die Urteile von Studierenden über die Lehrqualität an ihrer Hochschule lediglich Akzeptanzaussagen und keine Aussagen über die tatsächliche Lehrqualität darstellen (Kromrey 2000b, Kromrey 2000a, Kromrey 2001a).

- Die Bewertung muss wichtige Systemzusammenhänge, Abhängigkeiten und Rückkopplungsschleifen berücksichtigen und zwischen Eingangs-, Ausgangs- und Prozessgrößen trennen.
- Ein Kennzahlensystem muss eine geeignete Operationalisierung des Ranking-Ziels darstellen, d. h. es muss alle für die Ranking-Zielgruppe entscheidungsrelevanten Kriterien messen.
- Kennzahlen müssen unter Berücksichtigung von Sollwerten interpretiert werden.
- Die anschließend erforderliche Informationsverdichtung muss die Kennzahlenpräferenzen der Ranking-Zielgruppe repräsentieren.

Die Umsetzung einiger dieser Vorschläge erfordert umfangreiche statistische Erhebungen und Auswertungen. Durch den Einsatz neuer Publikationsmedien wie des Internet können diese jedoch in einem ersten Schritt „pragmatisch“ realisiert werden. Z. B. ist die Erstellung individueller Hochschul-Rankings mit einem Printmedium nicht möglich. Werden entscheidungsrelevante Daten jedoch auf einer CD-ROM oder im Internet potentiellen Nutzern (individuellen Ranking-Zielgruppen) zur Verfügung gestellt, so können diese individuelle Ranking-Objektmenge bilden, unter den Daten die für sie entscheidungsrelevanten auswählen, interpretieren und nach ihren individuellen Beurteilungsprofilen zu individuellen Hochschul-Rankings aggregieren.

Einige Ranking-Ersteller haben dies bereits erkannt und beginnen mit der Umsetzung. Die Studienführer von stern, start und CHE (1999,2000), das US News 2001 College Ranking und das Ranking von Business Week (2000) ermöglichen individuellen Ranking-Zielgruppen, individuelle Ranking-Objektmenge durch die Festlegung von Suchfiltern zu bilden. Business Week bietet online z. B. die zwei Suchfilter „Search by Statistics“ und „advanced MBA search“. Beim Suchfilter des Studienführers von stern, start und CHE (1999) können sieben Einzelfilter miteinander kombiniert werden; der Studienführer von stern, start und CHE (2000) ermöglicht die kombinierte Suche nach fünf Filtern, die bereits aus einer Menge von ca. 30 Einzelfiltern ausgewählt werden können. Die Suchkriterien und deren geforderte Werte bzw. Wertintervalle können in einem iterativen Prozeß so angepaßt werden, dass die resultierende Ranking-Objektmenge nur aus einer geringen Anzahl an Hochschulen besteht. In den beiden Extremfällen einer leeren und einer einelementigen Ranking-Objektmenge erübrigt sich ein Ranking.

Für den Fall einer mehrelementigen Ranking-Objektmenge können PC-gestützte, entscheidungsunterstützende Verfahren angewendet werden für die Festlegung individueller Zielsysteme, Kennzahlensysteme, Kennzahleninterpretationen und von Rankings, die individuelle Kennzahlenpräferenzen repräsentieren. Landeck (1994) entwickelte z. B. ein PC-Programm, das durch einfache Benutzerführung eine individuelle Zuordnung eines Studierenden zu einem von vier durch eine Cluster-Analyse bestimmten Lerntypen ermöglicht. Eine entsprechende Übertragung auf Klassen von Ranking-Zielgruppen z. B. innerhalb der Anspruchsgruppe der Studierberechtigten erscheint möglich. Hier kann die zielgruppenspezifische Strukturierung der Entscheidungssituation durch die Vorgabe zielgruppenspezifischer Zielsysteme, die Beschaffung aller entscheidungsnotwendigen Informationen durch zielgruppenspezifische Kennzahlensysteme, die zielgruppenspezifische Auswertung der Informationen durch zielgruppenspezifische Kennzahleninterpretationen und

die konsistente Verdichtung der Informationen zu einem Gesamturteil durch zielgruppenspezifische Rankings vorgegeben werden oder z. B. durch PC-gestützte Verfahren der multi-attributiven Entscheidungstheorie individuell bestimmt werden (von Nitzsch 1992).

Literatur

Arvidsson, G.. „Performance Evaluation“. *Guidance, Control, and Evaluation in the Public Sector*. Hrsg. F.-X. Kaufmann, G. Majone, V. Ostrum und W. Wirth. Berlin 1986. 625–43.

Baetge, J.. „Rating von Unternehmen anhand von Bilanzen“. *Die Wirtschaftsprüfung* 1 1994. 1–10.

Bankhofer, U. und A. Hilbert. „Eine empirische Untersuchung zum Berufseinstieg von Wirtschafts- und Sozialwissenschaftlern“. *Zeitschrift für Betriebswirtschaft* 65(12) 1995. 1423–1441.

Barz, A., P. Buhr, P. Giebisch, S. Hornbostel und D. Müller-Böling. „Studienführer 1999: Jura, Informatik, Mathematik, Physik - Vorgehensweise und Indikatoren“. *Arbeitspapier* 21. Centrum für Hochschulentwicklung (CHE), Gütersloh 1999.

Bathelt, H. und E. Giese. „Hochschul-Ranking als Instrument zur Belegung des Wettbewerbs zwischen Universitäten?“ *Münchener Geographische Hefte* 72 1995. 131–161.

Bayer, C. R.. „Spieltheoretische Analyse von Wettbewerbsstrategien in der westeuropäischen PKW-Reifenindustrie“. *Diplomarbeit* 1995. Universität Karlsruhe (TH), Institut für Wirtschaftstheorie und Operations Research.

Bayer, C. R.. *Hochschul-Ranking: Übersicht und Methodenkritik*. Beiträge zur Hochschulforschung. Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung. Sonderheft 1999. <<http://www.hochschul-ranking.de>>.

Bayer, C. R.. „Anspruch und Wirklichkeit von Hochschul-Ranking: Vorschlag einer allgemeinen Methodik“. *Die Betriebswirtschaft* 60(5) 2000. 547–69.

Bayer, C. R.. „Welchen Sinn haben Hochschul-Rankings und eindimensionale Signale im deutschen Hochschulsystem?“ *Die Betriebswirtschaft* 61(2) 2001. DBW-Dialog.

Bayer, C. R. und M. Krtscha. „Bewertung der Bilanzbonität im Rahmen der Jahresabschlußanalyse: Objektive Bewertungsfunktionen“. *Discussion Paper* 565 1999, Universität Karlsruhe (TH), Institut für Wirtschaftstheorie und Operations Research.

Bradburn, N. M.. „Das Ranking-Verfahren in den USA und seine Auswirkungen auf die Leistungsfähigkeit der Hochschulen“. *Vortrag zum 7. Bildungspolitischen Forum des Bundes Freiheit der Wissenschaft e.V.* 1987. The University of Chicago.

Brinkman, P. T. und D. J. Teeter. „Peer Institutional Studies/ Institutional Comparisons,“ *A Primer on Institutional Research*. Hrsg. J. A. Muffo und G. E. McLaughlin. Tallahassee 1987.

Buhr, P., P. Giebisch, S. Hornbostel und D. Müller-Böling. „Studienführer 2000: Architektur, Bauingenieurwesen, Elektrotechnik, Maschinenbau, Verfahrenstechnik, Physikalische Technik, Vermessungswesen, Wirtschaftsingenieurwesen - Vorgehensweise und Indikatoren“. *Arbeitspapier* 22. Centrum für Hochschulentwicklung (CHE), Gütersloh 2000.

Bundesministerium für Bildung und Forschung. *HRG – Hochschulrahmengesetz*. Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie. Bonn 1998.

- Campbell, D. T. „Reforms as experiments“. *American Psychologist* 24 1969. 409–29.
- CHE. „Projekte des Centrums für Hochschulforschung (CHE) im Bereich Strategie/Evaluation“. Online im Internet. URL: (<http://www.che.de/projekte.htm>) (Stand 99; Abfrage 2.11.99; MEZ: 14.26 Uhr). 1999.
- Chen, H. und P. H. Rossi. „The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science“. *Social Forces* 59 1980. 106–22.
- Chen, H. und P. H. Rossi. „Evaluation with sense: The theory-driven approach“. *Evaluation Review* 7 1983. 283–302.
- Chen, M. „Competitor analysis and interfirm rivalry: Toward a theoretical integration“. *Academy of Management Review* 21(1) 1996. 100–134.
- Cook, T. D. und D. T. Campbell. *Quasi-experimentation. Design & analysis issues for field setting*. Chicago 1979.
- Cook, T. D. und W. R. J. Shadish. „Program evaluation: The wordly science“. *Annual Review of Psychology* 37 1986. 193–232.
- Cronbach, L. J. *Designing evaluations of educational and social programs*. San Francisco 1982.
- Daniel, H.-D. „Die Wahl von Studienort und Hochschule“. *Effizient studieren*. Hrsg. E. Dichtl und M. Lingenfelder. Edition MLP, 3. überarbeitete und erweiterte Auflage. 3–22. Wiesbaden 1997.
- Daniel, H.-D. „Was bewirken Hochschul-Rankings“. *Hochschul-Ranking: Aussagefähigkeit, Methoden, Probleme*. Hrsg. D. Müller-Böling, S. Hornbostel und S. Berghoff. Gütersloh 2001. 123-126.
- Daniel, H.-D. und S. Hornbostel. „Die Studiensituation in der Betriebswirtschaftslehre im Urteil der Studenten“. *ZfB-Ergänzungsheft* 3 1993. 161–78.
- D'Aveni, R. A. „A multiple-constituency, status-based approach to interorganizational mobility of faculty and input-output competition among top business schools“. *Organization Science* 7(2) 1996. 166–189.
- Dettweiler, E. und H.-W. Frey. „Simulationsmodelle für die Hochschulplanung“. *HIS Brief* 19 1972. Hochschulinformationssystem GmbH. Hannover.
- Drew, D. E. und R. Karpf. „Ranking Academic Departments: Empirical Findings and a Theoretical Perspective“. *Research in Higher Education* 14(4). 1981.
- Eisenführ, F. und M. Weber. *Rationales Entscheiden*. 1994.
- Elsbach, K. D. und R. M. Kramer. „Member's Responses to Organizational Threats: Encountering and Countering the Business Week Rankings“. *Administrative Science Quarterly* 41 1996. 442–476.
- Fisch, R. „Ein Rahmenkonzept zur Evaluation universitärer Leistungen“. *Evaluation von Forschung: Methoden, Ergebnisse, Stellungnahmen*. Hrsg. H.-D. Daniel und R. Fisch. Band 4 aus der Serie Konstanzer Beiträge zur sozialwissenschaftlichen Forschung, 13–32. Konstanz 1988.
- Fischer, G., J. Rieker und S. Riesch. „Abschied von gestern“. *Manager Magazin* 2 1994. 138–171.

- Fischer, G. und U. Schwarzer. „Alma Marter“. *Manager Magazin* 1 1992. 148–163.
- Fischer, G. und U. Schwarzer. „Zurück in die Zukunft“. *Manager Magazin* 12 1992. 280–311.
- Focus. „Der große Uni-Test“. *Focus* 18 1997.
- Focus. „Deutschlands Top-Universitäten“. *Focus* 16 1997.
- Frackmann, E. „Die Bedeutung von ‚Rankings‘ für den Wettbewerb im Hochschulbereich“. *Betriebswirtschaftliche Forschung und Praxis* 2 1988. 139–162.
- Fürstenberg, F. „Kritische Würdigung von Hochschul-Rankings“. *Personal* 10 1995. 506–510.
- Gronwald, S. und T. Wöhrle. „Welche Hochschule bietet die besten Chancen für den Job?“. *Manager Magazin* 3 1998. 260–272.
- Guba, E. G. „Toward a methodology of naturalistic inquiry in educational evaluation“. *CSE Monograph Series in Evaluations*. Center for the Study of Evaluation, Los Angeles, California 1978.
- Guba, E. G. und Y. S. Lincoln. *Effective Evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco 1981.
- Guba, E. G. und Y. S. Lincoln. *Fourth Generation Evaluation*. Newbury Park, CA. 1989
- Haltmeier, H. „Aufruhr an den Unis: Der große Uni-Test“. *Manager Magazin* 25 1997. 146–151.
- Hamlén, W. und L. Southwick. „Quality in the MBA program: Inputs, outputs or value added?“. *Journal of Economic and Social Measurement* 15 1989. 1–26.
- Hammond, J. „Bringing Order into the Selection of a College“. *Personnel and Guidance Journal* 1965. 654–60.
- Hansen, U., T. Henning-Thurau und M. F. Langer. „Qualitätsmanagement von Hochschulen: FACULTY-Q als Erweiterung von TEACH-Q“. *Die Betriebswirtschaft* 60 2000. 23–38.
- Hansen, U., T. Henning-Thurau und H. Wochnowski. „TEACHQ: Ein valides und handhabbares Instrument zur Bewertung von Vorlesungen“. *Die Betriebswirtschaft* 57 1997. 376–96.
- Hess, W. „Wettbewerb und die Bewertung von Hochschulen“. *Beiträge zur Hochschulforschung*. 93–103, Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung 1992
- Himmelrath, A. „Hochschul-Rankings. Zahlenspiel mit Tücken“. *DUZ* 9 1997. 10–12.
- Hindmarsh, A., B. Kingston und R. Loynes. „League tables are getting better“. *Online im Internet* URL: (http://www.netnexus.org/ext/soapbox/leaguetales/thnexus_hindmarsh.html) 1999 (Stand 99; Abfrage 25.10.99; MEZ: 15.55 Uhr).
- Hodges, L. „Best for whom? University League Tables“. *The Times Higher Education Supplement*. 14. Mai 1993.
- Hornbostel, S. *Wissenschaftsindikatoren: Bewertungen in der Wissenschaft*. Opladen. 1997
- Hornbostel, S. „Der Studienführer des CHE - ein multidimensionales Ranking“. *Hochschul-Ranking. Zur Qualitätsbeurteilung von Studium und Lehre*. Hrsg. U. Engel. 2001.
- Huber, H. „Ein Kapazitätsmodell für Hochschulen“. *Saarbrücker Studien zur Hochschulentwicklung* 16, Hochschule des Saarlandes, Saarbrücken 1972.

- Jackson, S. „BW/ Harris Poll: How executives rate a business school education“. *Business Week*. 64. March 24th. 1986.
- Jensen, U. „Bootstrap-Rangkonfidenzintervalle für Hochschul-Ranglisten“. *Hochschul-Ranking: Aussagefähigkeit, Methoden, Probleme*. Hrsg. D. Müller-Böling, S. Hornbostel und S. Berghoff. Gütersloh 2001. 65-89.
- Kieser, A., E. Frese, D. Müller-Böling und N. Thom „Probleme der externen Evaluation wirtschaftswissenschaftlicher Studiengänge“. *Zeitschrift für Betriebswirtschaft*. ZfB-Ergänzungsheft 1/1996. 69–93.
- Klein, H. J. „Befragung neu-immatrikulierter Studierender an der Universität Karlsruhe (TH) zum Wintersemester 2000/01“. *Discussion Paper* Universität Karlsruhe (TH), Institut für Soziologie, Karlsruhe 2000.
- Kreutzmann, G. „Methodik Hochschul-Ranking - Fortschritt, aber noch offene Fragen“. *Die Betriebswirtschaft* 61 2001, 151–2, DBW-Dialog.
- Kromrey, H. „Evaluation der Lehre an der Ruhr-Universität Bochum“. *Discussion Paper*, Ruhr-Universität Bochum 1992.
- Kromrey, H. „Fallstricke bei der Implementations- und Wirkungsforschung sowie methodische Alternativen“. *Qualität von Humandienstleistungen. Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen*. Hrsg. H. Müller-Kohlenberg und K. Münstermann. 2000.
- Kromrey, H. „Qualität und Evaluation im System Hochschule“. *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Hrsg. R. Stockmann, Band 1 aus der Serie Sozialwissenschaftliche Evaluationsforschung. 233–58. 2000.
- Kromrey, H. „Studierendenbefragungen als Evaluation der Lehre? Anforderungen an Methodik und Design,“. *Hochschul-Ranking. Zur Qualitätsbeurteilung von Studium und Lehre*. Hrsg. U. Engel. 2001.
- Kromrey, H. „Zur Validität und zum Informationsgehalt von Studierendenbefragungen als Evaluation - Eine Reanalyse der CHE-Daten“. *Hochschul-Ranking: Aussagefähigkeit, Methoden, Probleme*. Hrsg. D. Müller-Böling, S. Hornbostel und S. Berghoff. Gütersloh. 2001. 43-63.
- Landeck, K.-J. *Lerntypdiagnose - Ein neues Verfahren für die Studienberatung*. Handbuch Hochschullehre, Kapitel E.2.2 1–17. Stuttgart 1994.
- Leitow, B. „Studentische Haltungen zur Studieneffizienz – ein Beitrag zur Typologie studentischer Orientierungen“. *Hefte zur Bildungs- und Hochschulforschung* 18. Konstanz 1996.
- Lewin, K., U. Heublein, J. Schreiber und D. Sommer. *Studienanfänger im Wintersemester 1998/99*. Band 138 Hochschulplanung. HIS Hochschul-Informationssystem GmbH. Hannover 1999
- MacDonald, B. „Evaluation and control of education“. *Evaluation: The state of the art*. Hrsg. D. Tawney. Schools Council, London 1975.
- Manager Magazin „Oben hui, unten pfui“. *Manager Magazin* 12 1990. 312–324.

- Miller, E. *Barron's Guide to Graduate Business Schools*. Barron's Educational Series, New York, 7 Auflage. 1990.
- Minke, E. und M. Weichhold „Kurzbericht zum Test des FIDES-Kapazitätsmodells“. *HIS Brief* 22, Hochschulinformationssystem GmbH, Hannover 1972.
- Peterson. *Graduate Schools in the U.S. 2001*. Princeton (NJ). 2000
- Porter, L. W. und L. E. McKibbin, *Management Education and Development: Drift or Thrust into the 21st Century?* New York 1988.
- Rieker, J. „Techno-Tempel“. *Manager Magazin* 3 1996. 236–247.
- Rieker, J. und S. Riesch. „Novizen aus der Provinz“. *Manager Magazin* 1 1995. 126–146.
- Rippey, R. M. (Hrsg.), *Studies in transactional evaluation*. Berkeley 1973.
- Rosigkeit, A. „Hochschul-Ranking: Hintergründe und kritische Anmerkungen zu einem modernen Bewertungsverfahren“. *Beiträge zur Hochschulforschung* 1 1997, Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung. 23–49.
- Segev, E., A. Raveh und M. Farjoun. „Conceptual Maps of the Leading MBA Programs in the United States: Core Courses, Concentration Areas, and the Ranking of the School“. *Strategic Management Journal* 20 1999. 549–565.
- Simon, H. „Positionierung von wirtschaftswissenschaftlichen Fachbereichen: Eine Grundlagenstudie zum strategischen Universitätsmarketing“. *Zeitschrift für Betriebswirtschaft* 55 1985. 827–847.
- Sinz, E. „Universitätsprozesse“. *Gestaltungskonzepte für Hochschulen: Effizienz, Effektivität, Evolution*. Hrsg. H.-U. Küpper und E. Sinz. 1998. 13–57.
- SMWK. *Gesetz über die Hochschulen im Freistaat Sachsen*. Staatsministerium für Wissenschaft und Kunst im Freistaat Sachsen. Köln, 11. Juni 1999.
- Solorzano, L. „America's best professional schools“. *U.S. News & World Report*, 70, November 2nd. 1987
- Sommer, C. „Juristic-Park“. *Manager Magazin* 2 1997. 138–151.
- Spiegel. „Die neuen Unis sind die besten“. *Der Spiegel* 50 1989. 70–87.
- Spiegel. „Willkommen im Labyrinth“. *Der Spiegel* 16 1993. 80–102.
- Spiegel. „Uni-Test Europa“. *Der Spiegel* 19 1998. 94–113, Teil I.
- Spiegel. „Uni-Test Europa“. *Der Spiegel* 20 1998. 72–91, Teil II.
- Spiegel. „Welche Uni ist die beste? Die Ergebnisse des Spiegel- Rankings“. *Der Spiegel* 15 1999.
- Stake, R. E. „The case study method in social inquiry“. *Educational Researcher* 7 1978. 5–8.
- Stern „93000 Noten von den Profs“. *stern* 16 1993. 172–184.
- Stern, Start und CHE. „Die Hochschulcharts: 566 Fachbereiche im Test“. *stern* 1999.
- Stern, Start und CHE. „Der Studienführer - Erste Hilfe für das Studium“. *stern* 2000.
- Stiftung Warentest. „Studienführer Chemie & Wirtschaftswissenschaften“. *Stiftung Warentest, test spezial* 98 01. 1998.

- Stufflebeam, D. L. „The CIPP model for program evaluation“. *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*. Hrsg. G. F. Madaus, M. S. Scriven und D. L. Stufflebeam. Boston 1983. 117–142.
- Stufflebeam, D. L. und A. J. Shinkfield. *Systematic Evaluation*. Boston, MA. 1985
- Syrbe, M. „Verlässliches Hochschul-Ranking: Eine Methodik aus einem Methoden-System für vielfältige Aufgaben“. *Die Betriebswirtschaft* 61 2001. DBW-Dialog. 148–51.
- Syrbe, M. und C. R. Bayer. „Ein quantitatives Wirkungsmodell von Universitäten“. *Beiträge zur Hochschulforschung*. 3 1997. Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung, 267–281.
- Tracy, J. und J. Waldfoegel. „The Best Business Schools: A Market- Based Approach“. *Journal of Business*. 70 1997. 1–31.
- Treinen, H. „Leistungsbewertung in der universitären Lehre“. *Discussion Paper*. Ruhr-Universität Bochum 1993.
- Turner, G. und G. Wiswede „Kriterien sind zu finden“. *Deutsche Universitätszeitung* 3 1986. 21.
- Turner, J. *Rediscovering the Social Group: A Self-categorization Theory*. Oxford 1987
- US News. „Business Methodology“. *Online im Internet*. 1998.
URL: (<http://www.usnews.com/usnews/edu/beyond/gradrank/gbbizmet.htm>) (Stand 98; Abfrage 25.1.99; MEZ: 9.03 Uhr).
- US News. „College rankings“. *Online im Internet*. 1998. URL:
(<http://www.usnews.com/usnews/edu/college/corank.htm>) (Stand 98; Abfrage 25.1.99; MEZ: 8.58 Uhr).
- US News. „Education Methodology“. *Online im Internet*. 1998. URL:
(<http://www.usnews.com/usnews/edu/beyond/gradrank/gbedmeth.htm>) (Stand 98; Abfrage 25.1.99; MEZ: 9.04 Uhr).
- US News. „Engeneering Methodology“. *Online im Internet*. 1998. URL:
(<http://www.usnews.com/usnews/edu/beyond/gradrank/gbengmet.htm>) (Stand 98; Abfrage 25.1.99; MEZ: 9.00 Uhr).
- US News. „Law Methodology“. *Online im Internet*. 1998. URL:
(<http://www.usnews.com/usnews/edu/beyond/gradrank/gblawmet.htm>) (Stand 98; Abfrage 25.1.99; MEZ: 9.02 Uhr).
- US News. „Medicine Methodology“. *Online im Internet*. 1998. URL:
(<http://www.usnews.com/usnews/edu/beyond/gradrank/gbmedmet.htm>) (Stand 98; Abfrage 25.1.99; MEZ: 9.08 Uhr).
- von Nitzsch, R. *Entscheidung bei Zielkonflikten: Ein PC-gestütztes Verfahren*. Band 95 aus der Serie neue betriebswirtschaftliche Forschung. Wiesbaden 1992, Zugleich: Technische Hochschule Aachen, Dissertation, 1991.
- von Trotha, K. „Leistungsprinzip in Staat und Gesellschaft, hier: Hochschulsystem“. *Landtag von Baden-Württemberg* 11. Wahlperiode, Drucksache 11/1555. 11. März 1993.
- Weiss, C. H. „Research for policy's sake: The enlightenment function of social research“. *Policy Analysis* 3 1977. 531–45.

Weiss, C. H. „Improving the linkage between social research and public policy“. *Knowledge and Policy: The Uncertain Connection*. Hrsg. L. E. Lynn. Washington D.C. 1978.

Welsch, H. und V. Ehrenheim „Ausbildung des wissenschaftlichen Nachwuchses: Zur Produktivität volkswirtschaftlicher Fachbereiche in Deutschland, Österreich und der Schweiz“. *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 119 1999. 455–73.

Westdeutsche Rektorenkonferenz. *Beurteilung und Entwicklung der Ansätze zur Leistungsbewertung und -messung von Hochschulen*. Westdeutsche Rektorenkonferenz, Bonn, Stellungnahme des 149. Plenums, 1.7.1986.

Westdeutsche Rektorenkonferenz. *Zum Wettbewerb im Hochschulbereich*. Westdeutsche Rektorenkonferenz, Bonn, Stellungnahme des 148. Plenums, 3./4.2.1986.

Wholey, J. S. *Evaluation and effective public management*. Boston MA. 1960

WZ „Die GhK in der Wahrnehmung ihrer neuen und ehemaligen Studierenden“. *WZ I update* 7, Wissenschaftliches Zentrum für Berufs- und Hochschulforschung, 34109 Kassel, <<http://www.uni-kassel.de/wz1/>>. 1999.

WZ „Abgeschlossen: Studierendenbefragungen an der GhK“. *WZ I update* 9, Wissenschaftliches Zentrum für Berufs- und Hochschulforschung, 34109 Kassel, <<http://www.uni-kassel.de/wz1/>>. 2000.

Was bewirken Hochschul-Rankings? Wer orientiert sich an ihnen?

Hans-Dieter Daniel

Vor gut zwei Jahren, am 12. April 1999, veröffentlichte der SPIEGEL zum dritten Mal Ranglisten der deutschen Universitäten. In den zwölf meiststudierten Fächern beurteilten mehr als 12000 Studierende die Studienbedingungen an ihren Universitäten. Jeweils zwei Professoren pro Fachbereich wurden von Interviewern des Emnid-Meinungsforschungsinstituts gefragt, welche Universitäten sie für das Studium ihres Faches empfehlen und von welchen sie abraten würden. In hunderten von Zeitungsartikeln und Leserbriefen wurden anschließend die Ergebnisse der Ranglisten diskutiert. Welchen Einfluss haben die Ranglisten auf die Hochschulvorlieben der Abiturienten? Wer sind die Abiturienten und Studienanfänger, die sich bei der Entscheidung für eine Universität an Rankings orientieren?

Über die Wirkung der Ranglisten bei Abiturienten ist bisher wenig bekannt. Welchen Einfluss haben das Urteil der Studierenden über die Studienbedingungen vor Ort einerseits und der Professorentipp andererseits auf die Studienort-Präferenz der Abiturienten? Für die fünf im SPIEGEL-Ranking berücksichtigten zulassungsbeschränkten Studiengänge Betriebswirtschaft, Rechtswissenschaft, Biologie, Psychologie und Medizin lässt sich die Frage vergleichsweise leicht beantworten, indem die Bewerberzahlen bei der Zentralstelle für die Vergabe von Studienplätzen (ZVS) in Dortmund vor Veröffentlichung der SPIEGEL-Ranglisten in Beziehung gesetzt werden zu den Bewerberzahlen nach der Veröffentlichung. Der Einfluss der Ranglisten auf die Studienplatznachfrage ist in allen fünf genannten ZVS-Studiengängen sehr ähnlich. Es besteht ein enger Zusammenhang zwischen der Platzierung einer Universität in der SPIEGEL-Rangliste nach dem Urteil der Studierenden („Wie zufrieden sind Sie alles in allem mit den Studienbedingungen in Ihrem Fach an dieser Universität?“) und der Zu- und Abnahme der Bewerberzahlen. Demgegenüber haben die Empfehlungen und Warnungen der Professoren („Welche drei Universitäten in Deutschland würden Sie für das Studium (Ihres Faches) empfehlen“ bzw. „Welche drei Universitäten würden Sie Ihrem Sohn oder Ihrer Tochter eher nicht empfehlen?“) keinen Einfluss auf die Hochschulpräferenz der Abiturienten.

Im Studiengang Medizin beispielsweise bewarben sich bei der ZVS für das Wintersemester 1999/2000 gegenüber dem Vorjahr umso mehr Abiturienten um einen Studienplatz, je besser die Studienbedingungen von den Studierenden vor Ort beurteilt wurden. Gegenüber dem Wintersemester 1998/1999 hat sich die Bewerberzahl an den besser beurteilten Universitäten um etwa 20 Prozentpunkte erhöht, an den schlechter beurteilten Universitäten ist sie dagegen um 20 Prozentpunkte zurückgegangen (die Korrelation zwischen dem Studierenden-Urteil und dem Quotienten aus der Bewerberzahl für das WS 1999/2000 und der Bewerberzahl für das WS 1998/1999 beträgt 0,57). Der Hochschultipp der Professoren wirkt sich dagegen kaum auf die Bewerberzahlen aus. Die Hochschulempfehlungen der Professoren sind mit 0,04 nahezu unkorreliert mit der Entwicklung der Studienplatznachfrage.

Wer sind die Abiturienten und Studienanfänger, die sich bei der Entscheidung für eine Universität an Rankings orientieren? Antworten auf diese Frage geben die bundesweite Studienanfängerbefragung der Hochschul-Informationssystem GmbH (HIS) in Hannover und die Befragung von Schülern der Jahrgangsstufe 13 an den Gymnasien des Haupteinzugsgebietes der Universität Gesamthochschule Kassel durch das Wissenschaftliche Zentrum für Berufs- und Hochschulforschung. An der Befragung der Studienanfänger des Wintersemesters 1998/1999, die HIS im Auftrag des damaligen Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie durchgeführt hat, beteiligten sich circa 5800 Studienanfänger. Den Fragebogen „Schule – und was dann?“ beantworteten im Frühjahr 1999 mehr als 1400 Abiturienten. In beiden Studien wurden die Studienanfänger bzw. Abiturienten gefragt, warum sie sich für ein Studium entschieden haben, welches die Gründe für die Wahl ihrer Hochschule und ihres Studienfaches waren und welches ihre Lebensziele sind.

Studienanfänger, die in der HIS-Studie angaben, dass bei der Wahl der Hochschule der Grund „Hochschule erhielt für mein Studienfach gute Bewertung in einer Hochschulrangliste/Ranking“ wichtig gewesen sei, unterscheiden sich deutlich von den Studienanfängern, die angaben, dieser Grund sei unwichtig bzw. unzutreffend gewesen. Der HIS-Studie zufolge werden Ranking-Listen am stärksten von Studienanfängern in den Wirtschafts-, Rechts- und Ingenieurwissenschaften sowie der Medizin zu Rate gezogen. Nur wenig Einfluss haben Rankings auf die Hochschulwahl von Studienanfängern in den Sozial-, Sprach- und Kulturwissenschaften. Männer orientieren sich häufiger an Rankings als Frauen.

Studienanfänger, denen es wichtig ist, an einer Hochschule zu studieren, die in einem Hochschul-Ranking gut bewertet wurde, studieren viel häufiger als Studienanfänger, denen dieser Grund unwichtig ist,

- um viele Berufsmöglichkeiten zu haben (65 % versus 47 %),
- um eine gesicherte Berufsposition zu erhalten (52 % versus 33 %),
- um einen angesehenen Beruf zu bekommen (38 % versus 22 %),
- um gute Verdienstmöglichkeiten zu erreichen (48 % versus 29 %).

Sie haben ihre Hochschule häufiger gewählt als Studienanfänger, denen die Platzierung ihrer Hochschule im Hochschul-Ranking unwichtig ist, wegen

- des guten Rufs von Hochschule und Professoren (73 % versus 13 %),
- der guten Ausstattung der Hochschule (44 % versus 14 %),
- des vielfältigen Lehrangebots (42 % versus 17 %),
- den überschaubaren Verhältnissen an der Hochschule (44 % versus 25 %).

Beide Gruppen von Studienanfängern unterscheiden sich auch deutlich in ihren Lebenszielen. Studienanfänger, denen es wichtig ist, an einer Hochschule zu studieren, die in einem Hochschul-

Ranking gut bewertet wurde, nennen viel häufiger als Studienanfänger, denen dieser Grund unwichtig ist, folgende Lebensziele:

- aus meinem Leben etwas machen, mich nicht treiben lassen (52 % versus 39 %),
- Erfolg haben (41 % versus 29 %),
- ein anerkannter Fachmann im Beruf werden (32 % versus 21 %).

Die Befragung der Kasseler Abiturienten im Frühjahr 1999 kommt zu sehr ähnlichen Ergebnissen wie die Studienanfängerbefragung der HIS GmbH. Die Studie des Wissenschaftlichen Zentrums für Berufs- und Hochschulforschung zeigt neben den bereits im Zusammenhang mit der HIS-Studie berichteten Gruppenunterschieden, dass Abiturienten, die sich bei der Wahl ihrer Wunschhochschule an Rankings orientieren, stärker als die übrigen Abiturienten an einer wissenschaftlichen Ausbildung (84 % versus 72 %) und an Forschung (51 % versus 39 %) interessiert sind. Sie streben häufiger eine leitende berufliche Funktion an (97 % versus 83 %), möchten häufiger eine Arbeit haben, die zum technischen Fortschritt beiträgt (43 % versus 32 %) und in fachlicher Hinsicht Überdurchschnittliches leisten (88 % versus 77 %). Diejenigen, die sich bei der Hochschulwahl nicht an Rankings orientieren, studieren dagegen häufiger, weil sie vor bzw. nach der Routine des Berufslebens noch etwas anderes erleben möchten (67 % versus 57 %); sie möchten in ihrem späteren Berufsleben häufiger als Abiturienten, die sich an Rankings orientieren, zu sozialen Veränderungen beitragen (63 % versus 50 %), kreativ-gestalterisch tätig sein (56 % versus 46 %) und sich für den Umweltschutz einsetzen (42% versus 34 %).

Resümee

Die vorliegenden Studien zeigen, dass das häufigste Entscheidungskriterium bei der Wahl des Hochschulortes nicht der Rang der Hochschule, sondern nach wie vor die regionale Nähe zum Heimatort ist (vgl. Muske, 1975; Peisert et al., 1983; Stein, 1977). Gleichwohl gibt bei bundesweiten Befragungen ein Viertel der Studienanfänger an, dass der gute Ruf der Hochschule ein „Hochschulwahlmotiv mit großer Bedeutung„ gewesen sei (Studienanfänger insgesamt: 26 %, Studienanfänger an Universitäten: 28 %, Studienanfänger an Fachhochschulen: 21 %; vgl. Lewin et al., 2000, S. 85). Eine Auswertung der Bewerberzahlen bei der Zentralstelle für die Vergabe von Studienplätzen (ZVS) vor und nach der Veröffentlichung von Hochschulranglisten zeigt, dass an den besser beurteilten Universitäten sich die Bewerberzahl um etwa 20 Prozent erhöht, an den schlechter beurteilten Universitäten dagegen um etwa 20 Prozent zurückgeht. Studienberechtigte und Studienanfänger, die sich bei der Wahl der Hochschule an Ranglisten orientieren, sind nach den vorliegenden Ergebnissen ehrgeiziger und karriereorientierter als nicht an Rankings interessierte Abiturienten und Kommilitonen.

Literaturhinweise

Daniel, Hans-Dieter & Münch, Katrin „Erfahrungen aus den Evaluationsprojekten in Kassel und Mannheim“. *Leitbild der Hochschule – Qualität der Lehre*. Hrsg. Hochschulrektorenkonferenz. 97-114. Bonn: HRK (Beiträge zur Hochschulpolitik 2/2000).

Lewin, Karl, Heublein, Ulrich, Schreiber, Jochen & Sommer, Dieter. *Studienanfänger im Wintersemester 1998/99*. Hannover: HIS GmbH (Hochschulplanung 138). 1999

Muske, Gitta. *Motive für die Wahl des Studienortes München*. Kallmünz/Regensburg. 1975.

Peisert, Hansgert, Framheim, Gerhild, Kuthe, Manfred & Reinhardt, Klaus. *Alte und neue Universitäten – Einzugsbereiche und Ortswahl der Studenten, Motive und Verhalten*. Bonn: Der Bundesminister für Bildung und Wissenschaft (Schriftenreihe Hochschule 44). 1983

Stein, Wolfgang. *Regionale und soziale Herkunft der Studierenden – Zugänge zur Universität Erlangen-Nürnberg*. Frankfurt/Main 1977.

Drittmittel als Indikator für Forschungsleistung.

Analysen zum DFG-Bewilligungsaufkommen

Jürgen Güdler

Zusammenfassung

Im Juni 2000 veröffentlichte die Deutsche Forschungsgemeinschaft (DFG) zum zweiten Mal eine nach Hochschulen und Fachgebieten differenzierte Übersicht ihrer Bewilligungen, ein erster Bericht war 1997 erschienen. Der Beitrag setzt sich zunächst mit den Zielsetzungen auseinander, die mit diesen so genannten „DFG-Rankings“ verbunden sind. Anschließend werden disziplinäre Unterschiede im DFG-Drittmittelbedarf herausgearbeitet. Der Hauptteil des Beitrags ist der Diskussion methodischer Fragen gewidmet. Im Vordergrund stehen dabei vor allem Aspekte der fachlichen Differenzierung sowie der Relativierung zur Zahl der an einer Hochschule tätigen Wissenschaftler.

1. Hintergrund und Zielsetzungen

Die Deutsche Forschungsgemeinschaft ist der größte Drittmittelgeber für grundlagenorientierte Forschung in Deutschland. Im Jahr 1999 wurden etwa 2,4 Milliarden DM an Fördermitteln bewilligt, davon 88,6 Prozent für Forschung an Hochschulen. Eine Erhebung des Wissenschaftsrats weist für das Jahr 1997 einen Anteil der DFG-Mittel an Drittmiteleinwerbungen der Hochschulen von 35 Prozent aus (Wissenschaftsrat 2000: 59f).

Im Herbst 1996 entsprach die Deutsche Forschungsgemeinschaft dem Wunsch einer Gruppe von Rektoren und Präsidenten großer Hochschulen nach Angaben über die zehn - gemessen am gesamten DFG-Fördervolumen - erfolgreichsten Universitäten. Der öffentlichen Bekanntgabe dieser „Top-Ten“-Zahlen folgte eine lebhafte Diskussion, die schließlich in der Forderung mündete, eine breitere Datenbasis bereitzustellen. Die Deutsche Forschungsgemeinschaft kam dieser Forderung nach und veröffentlichte 1997 erstmals einen Bericht, der umfassend Auskunft über die Bewilligungsvolumina deutscher Universitäten der Jahre 1991 bis 1995 gab.

Die Veröffentlichung stieß auf breite Resonanz. Der gedruckte Bericht wurde in mehr als dreitausend Exemplaren bei der Geschäftsstelle angefordert, eine nachträglich über das Internet zur Verfügung gestellte Onlineversion ging in ähnlicher Stückzahl über den „elektronischen Ladentisch“. Aber nicht nur der quantitative Umfang der Nachfrage sowie der Medienberichterstattung über diese Dokumentation, auch die qualitative Nutzung der dort präsentierten Daten war beeindruckend. Erfolgte bis dahin die Offenlegung von Zahlen in erster Linie aus Gründen der Rechenschaftspflicht gegenüber den Geldgebern, sah sich die Geschäftsstelle der DFG plötzlich mit neuen Zielgruppen konfrontiert: Hochschulen und mit Forschungsförderung befasste Bundes- und Landesministerien verwendeten die Zahlen als Hintergrundinformation für Forschungsplanung - auch im Wettbewerb um immer knapper werdende Ressourcen. (Nachwuchs-)Wissenschaftler aus dem In- und Ausland nutzen die Daten, um sich über die Zentren der Forschung in Deutschland zu informieren. Mit der Veröffentlichung hat die DFG einen Beitrag zur Diskussion um universitäre

Forschungsevaluation geleistet. Dass die Relevanz entsprechender Studien zunimmt, zeigt die Vielzahl an Veröffentlichungen, die - etwa durch das Centrum für Hochschulentwicklung (CHE) - zu diesem Thema in den letzten Jahren vorgelegt wurden. Die Medien schließlich sahen in den von der DFG offengelegten Zahlen einen willkommenen Beitrag zu dem Mitte der 90er Jahre einsetzenden Wettstreit um die aussagekräftigsten „Hitlisten“ und „Hochschul-Rankings“.

Die breite Resonanz war für die Geschäftsstelle der DFG Anlass genug, das sog. „DFG-Ranking“ zu einem festen Bestandteil ihres Informationsportfolios zu machen. Eine aktuelle Bewilligungsübersicht für die Jahre 1996 bis 1998 wurde im Juni 2000 vorgelegt¹. In Erweiterung zum ersten Bericht gingen hier auch Zahlen zu außeruniversitären Forschungseinrichtungen ein. Zudem erfolgte eine stärker die fachlichen Schwerpunktsetzungen von Universitäten herausarbeitende Aufschlüsselung der Daten. Der Bericht wurde schließlich auch genutzt, um in einem einleitenden Hintergrundkapitel über die Angebots- und Nachfrageentwicklung bei von der DFG angebotenen Förderprogrammen und über erste Konsequenzen zu informieren, die aus einer 1998 durchgeführten Systemevaluation von Deutscher Forschungsgemeinschaft und Max-Planck-Gesellschaft (MPG) durch eine internationale Expertenkommission² gezogen wurden.

2. Die DFG im Kontext anderer Förderer

Die DFG ist der größte Forschungsförderer für Forschung an Universitäten - gleichwohl ist sie nicht die einzige Einrichtung, die Forschung an Hochschulen finanziert. Das so genannte „DFG-Ranking“ ist deshalb kein „Drittmittel-Ranking“ (in dem Sinne, das es hinreichend Aufschluss gibt über die gesamte Akquiseleistung einzelner Hochschulen), vor allem aber ist es kein „Forschungs-Ranking“, das - gestützt auf einen breiten Set an Indikatoren - umfassend Auskunft über die Forschungsleistung einzelner Hochschulen bietet. Allein das Mittelvolumen, das durch die DFG für bestimmte Fachgebiete zur Verfügung gestellt beziehungsweise bewilligt wurde, bildet die Analysegrundlage.

„Drittmittel-Rankings“ oder gar „Forschungs-Rankings“ können nur dann objektiv Auskunft über die „Leistungsfähigkeit“ einer Hochschule geben, wenn sie zwei Dinge berücksichtigen:

- Leistungsvergleiche sollten nicht über klar definierte Fachgrenzen hinaus erfolgen und
- Leistungsvergleiche sollten sich auf eine möglichst große Zahl an Indikatoren stützen - DFG-Drittmittel bilden hier nur einen (wenn auch gewichtigen) Teil eines sorgfältig zu bestimmenden Messinstrumentariums.

¹ Der Bericht „DFG-Bewilligungen an Hochschulen und außeruniversitäre Forschungseinrichtungen - 1996 bis 1998“ kann beim Pressereferat der DFG angefordert werden, eine Online-Version sowie kartographische Darstellungen der regionalen Verteilung von DFG-Bewilligungen sind unter <http://www.dfg.de/berichtswesen/ranking.html> abrufbar.

² Der Bericht der Kommission ist unter <http://www.blk-bonn.de/papers/forschungsfoerderung.pdf> abrufbar. Eine gemeinsame Stellungnahme von DFG und MPG findet sich unter <http://www.dfg.de/aktuell/download/evaluation.html>.

Der quantitativ hohe Anteil, den DFG-Drittmittel an Drittmittelinwerbungen universitärer Einrichtungen einnehmen, und sicher auch die besondere Wertschätzung, die Fördermittel der DFG in der wissenschaftlichen Gemeinschaft genießen, tragen gleichwohl dazu bei, daß statistischen Übersichten, die DFG-Bewilligungen nach Hochschulen und Fachgebieten differenziert ausweisen, eine Aussagekraft eigenen Wertes zuzusprechen ist. Die Entscheidung, ein Projekt mit Mitteln der DFG zu finanzieren, erfolgt im Rahmen strenger Auswahlverfahren durch Peers - eine große Zahl DFG-geförderter Projekte und Programme an einer Hochschule (und dort: in einem bestimmten Fachgebiet) weist darauf hin, dass die Forschung der dort tätigen Wissenschaftler in der scientific community große Anerkennung genießt. Aufmerksamkeit gewinnen die von der DFG vorgelegten Zahlen auch deshalb, weil sie nahezu „konkurrenzlos“ sind: Keine andere Fördereinrichtung in Deutschland hat ihre Förderleistungen bisher in ähnlich differenzierter Form offengelegt, keine Organisation hat sich bisher darum bemüht, ein regelmäßiges Berichtswesen zu implementieren, das für einen breiten Fächer- und Disziplinenkanon Daten zu Drittmittelinwerbungen und anderen Forschungsindikatoren bereitstellt.

Einen wichtigen Beitrag leistet in diesem Zusammenhang das Centrum für Hochschulentwicklung (CHE), das in seinen Studienführern für ausgewählte Fächer unter anderem eine Reihe an Kennzahlen präsentiert, die Auskunft auch über vor Ort erbrachte Forschungsleistungen bieten. Der in diesem Band veröffentlichte Beitrag von Stefan Hornbostel offenbart etwa am Beispiel von Drittmittelinwerbungen das breite Spektrum, das von Wissenschaftlern zur Finanzierung ihrer Forschung in Anspruch genommen wird. Beeindruckend sind vor allem die Unterschiede, die sich zwischen den Fächern - etwa den Rechts- und den Ingenieurwissenschaften - zeigen. Während auf der einen Seite weniger als 20 Prozent aller eingeworbenen Drittmittel von der DFG stammen, sind es auf der anderen Seite nahezu 50 Prozent (vgl. Abbildung 13 im Beitrag von Stefan Hornbostel, S. 32 in diesem Band). Auch der generelle Stellenwert von Drittmitteln ist stark von fachkulturellen Besonderheiten geprägt. In manchen Fächern sind Drittmittel ein unverzichtbares Instrument zur Durchführung apparativ und/oder personell aufwendiger Forschung (etwa der Biologie), in anderen Fächern (etwa der Mathematik) lässt sich ein großer Teil der Forschungsleistungen auch ohne umfangreiche Drittmittel bewältigen.

Eine Antragstellerbefragung, die die DFG aus Anlass ihrer Evaluation 1998 durchführte, kommt zu ähnlichen Ergebnissen. Die Erhebung diente dem Ziel, Informationen über die allgemeinen Forschungsaktivitäten der DFG-Klientel, über ihren Bedarf an Drittmitteln, ihre Einstellung zum Gutachtersystem, ihre Zufriedenheit mit dem Förderangebot und - nicht zuletzt - ihre Zufriedenheit mit der Arbeit der Geschäftsstelle zu ermitteln. Befragt wurden Wissenschaftler, die in den Jahren 1992 bis 1996 mit Anträgen für Projekte oder Stipendien an die DFG herangetreten waren. In der postalischen Erhebung, die vom Mannheimer Zentrum für Umfragen, Methoden und Analysen (ZUMA) durchgeführt wurde, konnten 1.256 Antragsteller befragt werden³.

Eine der in dieser Erhebung gestellten Fragen richtete sich auf die Forschungs- oder besser: Förderkontexte, in denen Antragsteller der DFG aktiv sind. Mit der Frage:

³ Ergebnisse der Befragung liegen bisher nur in Form DFG-interner Arbeitsberichte vor. Die Veröffentlichung eines Abschlussberichts ist vorgesehen.

Welche der folgenden Projektarten und Beratungstätigkeiten haben Sie in den letzten fünf Jahren (federführend) durchgeführt?

wurden sie um Auskunft zu ihrer weiteren drittmittelfinanzierten Forschungstätigkeit gebeten. Die Ergebnisse weisen darauf hin, dass für große Teile der Befragten DFG-Drittmittel einen zwar zentralen aber keineswegs ausschließlichen Bestandteil ihres Fördermittelkatalogs bilden.

Wichtigster Kooperationspartner der Antragsteller der DFG ist das Bundesministerium für Bildung und Wissenschaft, Forschung und Technologie (BMBF). Jeder vierte Antragsteller gibt an, in den letzten fünf Jahren mit Mitteln des BMBF geforscht zu haben. Von großer Bedeutung sind aber auch die Fördereinrichtungen der Länder sowie Stiftungen (je 24 Prozent). Ein nicht unbeträchtlicher Anteil der Befragten ist überdies in internationale Kontexte eingebunden: Die Förderprogramme der Europäischen Union werden von jedem fünften Befragten genutzt, jeder zehnte DFG-Antragsteller kann auf Erfahrungen mit Fördereinrichtungen im Ausland zurückblicken (vgl. Abbildung 1).

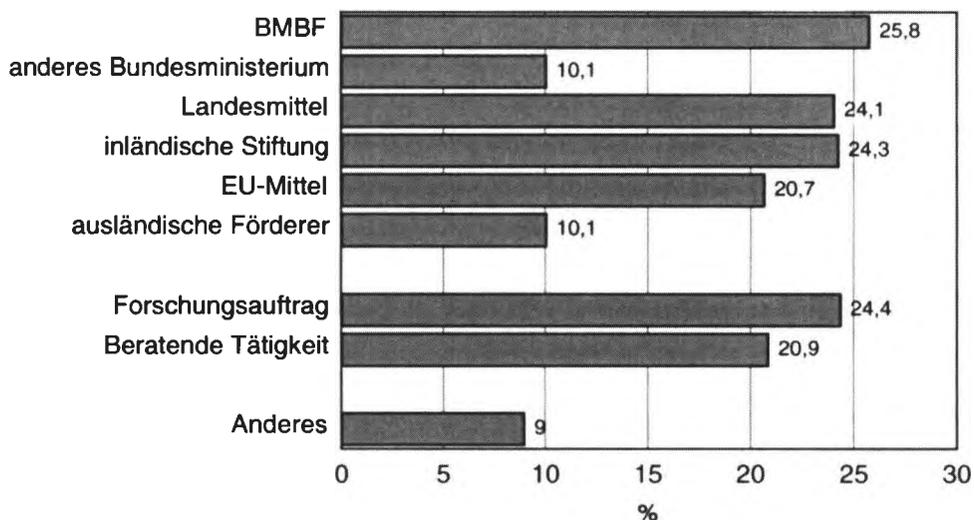


Abbildung 1: Fördereinrichtungen und Projektarten, die von Antragstellern der DFG innerhalb eines 5-Jahre-Zeitraums (1992-1996) genutzt wurden (in Prozent der befragten Antragsteller)

Eine zusätzliche Dimension der wissenschaftlichen Tätigkeit sollte mit den Antwortvorgaben „Forschungsauftrag einer staatlichen Instanz, eines Unternehmens oder eines Verbandes“ und „Beratende Tätigkeit für eine staatliche Instanz, ein Unternehmen oder einen Verband“ abgebildet werden. Hier steht nicht die finanzierende Institution im Vordergrund, sondern die Art der Beziehung, in der ein Forscher zu einer Einrichtung steht. Fast ein Viertel der DFG-Antragsteller hat in den letzten fünf Jahren Auftragsforschung betrieben, und rund 20 Prozent waren beratend für eine Einrichtung tätig. DFG-finanzierte Forscher tragen damit in erheblichem Umfang zum Wissenstransfer bei.

Insgesamt wird deutlich, daß die DFG in ein Netzwerk von Drittmittelgebern eingebunden ist. DFG-Antragsteller akquirieren in nennenswertem Umfang Fördermittel aus unterschiedlichen Quellen: Zwei von drei Antragstellern der DFG kooperieren mit mindestens einer weiteren Fördereinrichtung, 17 Prozent der Befragten haben sogar Mittel von drei und mehr Einrichtungen erhalten (vgl. Abbildung 2).

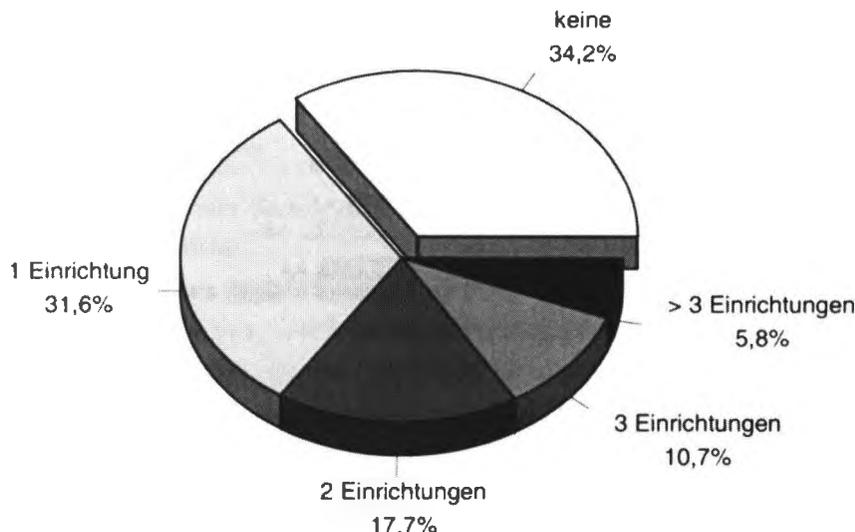


Abbildung 2: Zwischen 1992 und 1996 zusätzlich von DFG-Antragstellern in Anspruch genommene Förderer (in Prozent)

Die DFG bedient damit mehrheitlich keine in sich geschlossene Klientel, ihre Mittel werden vielmehr von Wissenschaftlern genutzt, die zur Durchführung ihrer Forschungsarbeiten ergänzend auch auf Mittel anderer Förderer zugreifen.

3. Fachliche Unterschiede im DFG-Drittmittelbedarf

Es ist kaum überraschend, daß in Abhängigkeit von der Fachzugehörigkeit erhebliche Unterschiede in der Nutzung weiterer Drittmittelquellen bestehen. Ingenieurwissenschaftler etwa nehmen sämtliche Fördereinrichtungen, namentlich das BMBF und die Förderinstanzen der Länder, in weit überdurchschnittlichem Umfang in Anspruch; Antragsteller aus diesem Bereich treten überdies doppelt so häufig als Auftragsforscher und Berater in Erscheinung wie der Durchschnitt aller Befragten. Fachliche Unterschiede lassen sich aber auch anhand weiterer Statements der Antragstellerbefragung herausarbeiten.

Die folgende Frage diente der Ermittlung der allgemeinen Antragsaktivität:

Wenn Sie alle Anträge, die Sie in den letzten fünf Jahren bei Zuwendungsgebern eingereicht haben, zusammennehmen: Wie viele Anträge haben Sie bei Zuwendungsgebern insgesamt eingereicht, und wie viele dieser Anträge sind bewilligt worden?

In Bezug auf DFG-Anträge wurde eine ähnlich lautende Frage gestellt:

Wie viele Anträge - gegliedert nach Antragsformen⁴ - haben Sie in den letzten fünf Jahren an die DFG gerichtet, und wieviele davon sind bewilligt worden (im Zweifel bitte circa-Angaben)?

Allgemein haben DFG-Antragsteller im Schnitt etwas mehr als sieben Anträge bei Förder- einrichtungen eingereicht, von denen fast fünf bewilligt wurden (vgl. Abbildung 3).

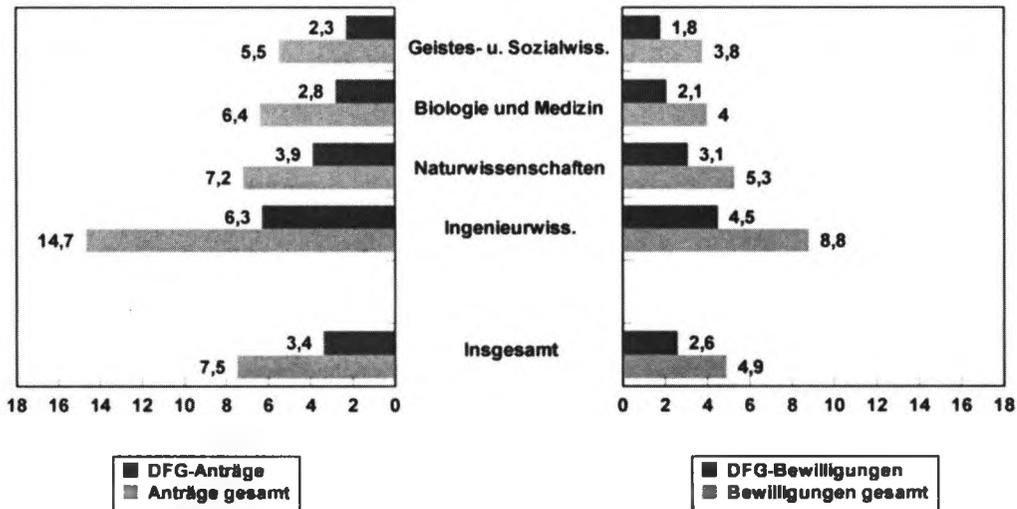


Abbildung 3: Mittlere Zahl der zwischen 1992 und 1996 eingereichten und bewilligten Anträge bei Drittmittelgebern insgesamt sowie bei der DFG je DFG-Antragsteller

In den vier Wissenschaftsbereichen liegt der Anteil, den DFG-Anträge am gesamten Antrags- wie auch Bewilligungsaufkommen einnehmen, in einem relativ engen Spektrum zwischen 42 und 54 (Anträge), beziehungsweise 47 und 58 Prozent (Bewilligungen). Relativ das höchste Gewicht nimmt die DFG in den Naturwissenschaften ein. DFG-Antragsteller, das zeigt sich auch hier, nutzen fachübergreifend in relevantem Umfang weitere Quellen, um ihre Forschungsaktivitäten zu finanzieren.

Größere Unterschiede ergeben sich hinsichtlich der absoluten Drittmittelaktivität befragter Wissenschaftler. Auffallend ist auch hier wiederum der ingenieurwissenschaftliche Bereich. Dort aktive DFG-Antragsteller reichen pro Person bei der DFG überdurchschnittlich viele Anträge ein, nutzen aber auch andere Finanzierungsquellen in weit größerem Umfang als Antragsteller aus anderen Wissenschaftsbereichen. Entsprechend ihrer erhöhten Antragsaktivität erhalten Ingenieurwissenschaftler in ebenfalls überdurchschnittlichem Umfang Bewilligungen. Sie werben Mittel für knapp neun Projekte in fünf Jahren ein - der allgemeine Mittelwert liegt bei fünf Projekten.

Die eben angedeuteten fachlichen Unterschiede in der Nutzung verschiedener Drittmittelquellen zeigen sich schließlich auch im Umfang der Mittel, die je Fachgebiet bei der DFG eingeworben werden. Die folgende Tabelle, die der aktuellen Ausgabe des „DFG-Rankings“ entnommen wurde, stellt die in den Jahren 1996 bis 1998 bewilligten Mittel in insgesamt 16 Fachgebieten

⁴ Als Antwortkategorien waren die wichtigsten Förderprogramme der DFG aufgeführt (Einzelförderung, Forschergruppen, Teilprojekte zu Sonderforschungsbereichen etc.).

der Zahl der Professoren und Wissenschaftler insgesamt gegenüber, die in entsprechenden Lehr- und Forschungsgebieten tätig waren.

	Mio DM	Professoren		Wissenschaftler insg.	
		N	TDM je Professor	N	TDM je Wiss.
Gesellschaftswissenschaften	213,3	3.158	67,5	12.398	17,2
Geschichts- und Kunstwissenschaften	200,3	1.163	172,3	3.577	56,0
Sprach- und Literaturwissenschaften	176,4	1.922	91,8	7.913	22,3
Psychologie, Pädagogik, Philosophie, Theologie	183,6	1.995	92,0	6.800	27,0
Geistes- und Sozialwissenschaften	773,6	8.238	93,9	30.688	25,2
Medizin	920,0	3.137	293,3	37.724	24,4
Biologie	780,4	963	810,4	5.847	133,5
Veterinärmedizin	15,9	196	81,4	839	19,0
Agrar- und Forstwissenschaften	121,6	531	229,0	3.234	37,6
Biologie / Medizin	1.838,0	4.827	380,8	47.644	38,6
Geowissenschaften	246,5	422	584,1	2.334	105,6
Chemie	406,6	1.091	372,7	9.107	44,6
Physik	523,6	1.182	443,0	7.741	67,6
Mathematik	133,2	1.214	109,8	4.003	33,3
Naturwissenschaften	1.309,9	3.909	335,1	23.185	56,5
Allg. Ingenieurwissenschaften und Maschinenwesen	837,0	1.025	816,6	9.160	91,4
Architektur, Städtebau, Bauingenieurwesen	71,8	911	78,8	4.971	14,4
Bergbau und Hüttenwesen	68,3	74	923,3	532	128,4
Elektrotechnik, Informatik	347,2	1.160	299,3	7.814	44,4
Ingenieurwissenschaften	1.324,2	3.170	417,7	22.477	58,9
Insgesamt	5.245,7	20.144	260,4	123.994	42,3

Quelle Personenzahlen: Statistisches Bundesamt, Hauptberuflich tätiges wissenschaftliches und künstlerisches Personal an Hochschulen (Stand: 1998). Sonderauswertung im Auftrag von DFG und HRK (Basis: 72 Universitäten, die zwischen 1996 und 1998 mehr als 3 Millionen DM DFG-Bewilligungen erhalten haben).

Tabelle 1: DFG-Bewilligungen 1996 bis 1998 im Verhältnis zur Zahl der Professoren/Wissenschaftler insgesamt je Fachgebiet

Zu erkennen ist, daß DFG-Mittel in ganz unterschiedlichem Umfang zur Finanzierung von Forschung genutzt werden. Insgesamt entfallen auf mehr als 20.000 Professoren an den für diese Gegenüberstellung berücksichtigten 72 Hochschulen 5,2 Milliarden DM DFG-Fördermittel in drei Jahren. Dies entspricht einem durchschnittlichen Bewilligungsvolumen von 260.000 DM je hauptberuflich tätigem Professor. In Relation zur Zahl hauptberuflich tätiger Wissenschaftler insgesamt (124.000 Personen) ergibt sich ein Durchschnittswert von 42.000 DM. Zwischen den Fachgebieten zeigen sich allerdings beachtliche Unterschiede. Ins Auge fallen vor allem die Allgemeinen Ingenieurwissenschaften (einschl. Maschinenwesen), das Fachgebiet Bergbau und Hüttenwesen und die Biologie. In diesen drei Gebieten wurden je Professor zwischen 810.000 und 920.000 DM bewilligt. Vergleichsweise gering ist dagegen das auf Personen relativierte Bewilligungsaufkommen in der Medizin. In diesem Fachgebiet, das absolut mit weitem Abstand den größten Anteil am

DFG-Bewilligungsvolumen eingeworben hat, entfallen auf einen Professor nur etwas weniger als 300.000 DM in drei Jahren.

Die geringste Inanspruchnahme ergibt sich für Fachgebiete des geistes- und sozialwissenschaftlichen Spektrums. Dort sind - mit Ausnahme der Geschichts- und Kunstwissenschaften - Durchschnittsbeträge unter 100.000 DM typisch. Vor allem die Gesellschaftswissenschaften, also Fächer wie Soziologie und Geographie, Jura und Wirtschaftswissenschaften, stehen der DFG vergleichsweise fern - ohne dass sich hieraus der Schluß ziehen ließe, diese Fächer seien relativ forschungspassiv.

4. Methodische Aspekte

4.1 Datenbasis

Das „DFG-Ranking“ ist Teil eines Informationsportfolios, dessen gemeinsame Grundlage die Förderdatenbanken der Deutschen Forschungsgemeinschaft bilden. Diese Datenbanken dienen in erster Linie der Unterstützung interner Geschäftsabläufe. Sie werden für die halbautomatische Erstellung von Eingangsbestätigungsschreiben zu Förderanträgen, für die Erstellung von Briefen an (potentielle) Gutachter, für die Generierung von Antragsübersichten, die der Information von Entscheidungsgremien der DFG dienen und schließlich auch für die Abfassung standardisierter Bewilligungs- und Ablehnungsschreiben genutzt.

Weiterhin dienen die Datenbanken der DFG als Grundlage für statistische Übersichten. Diese werden erstellt, um vor allem geschäftsstellenintern einen regelmäßigen Überblick über die Entwicklung des Antragsengagements sowie der Bearbeitung von Anträgen zu gewinnen. Zunehmende Bedeutung gewinnen diese Quellen dabei auch als Basis für internes Controlling. So lassen sich etwa Aussagen treffen über die Entwicklung der Antragsbelastung einzelner Fachreferate oder zur Bearbeitungsdauer, die für dort bearbeitete Anträge benötigt wird. Die Offenlegung entsprechender Zahlen trägt dazu bei, sowohl Überbelastungen als auch Schwachstellen zu erkennen und geeignete Maßnahmen zu deren Beseitigung einzuleiten.

4.2 Institutionelle Zuordnung von DFG-Bewilligungen

Die Datenbanken der DFG archivieren sogenannte „prozeßproduzierte“ Daten. Der Erstellungszweck dieser Daten ist primär nicht, Grundlagen für evaluative Studien bereitzustellen, sondern - wie an Beispielen ausgeführt - Geschäftsabläufe zu unterstützen. Am Beispiel des „DFG-Rankings“ zeigt sich dieses Defizit vor allem hinsichtlich der fehlenden Möglichkeit, Aussagen zu Bewilligungen in nach Empfängerinstituten differenzierter Form treffen zu können. „Im Mittelpunkt steht der Antrag“ - diese Maxime, die - lange Zeit ausschließlich - das Handeln der Geschäftsstelle der Deutschen Forschungsgemeinschaft prägte, spiegelt sich in gewisser Weise auch im Aufbau der DFG-Datenbanken wider: In detaillierter Form werden dort Informationen über die mit einem Antrag beantragten und bewilligten Mittel, differenziert nach sog. „Kostenarten“ (wiss. Personal, studentische und nichtwissenschaftliche Mitarbeiter, Geräte, Apparaturen, Reisemittel etc.) aufgeschlüsselt, werden Informationen zu den „Verfahren“ (etwa: Normalverfahren, Schwerpunktprogramm) und „Beihilfearten“ (Sachbeihilfen, Druckbeihilfen,

Stipendien etc.) dokumentiert, in deren Rahmen diese Fördermaßnahmen finanziert werden. Schon die Daten zur Person des Antragstellers dagegen sind weniger differenziert, erst seit wenigen Jahren wird etwa mit Nachdruck auf die Erfassung des Geburtsdatums eines Antragstellers geachtet.

Vor allem aber - im Kontext der Ranking-Diskussion von besonderem Belang - erfolgen statistisch analysierbare Angaben zur institutionellen Herkunft eines Antragstellers nach wie vor nur in sehr hoch aggregierender Form - mit Erfassung eines sog. „Forschungstellenschlüssels“, der über die Einrichtung informiert, an der ein Wissenschaftler zum Zeitpunkt der Antragstellung tätig ist. Auf diese Weise ist es zwar möglich, eine Aussage über den Namen der Hochschule oder der außeruniversitären Forschungseinrichtung zu treffen, aus der ein Antrag stammt; eine Aufgliederung nach Fachbereichen, Fakultäten und Instituten lässt sich dagegen nicht vornehmen.

4.3 Fachliche Zuordnung von DFG-Bewilligungen

Fachlich differenzierende Aussagen zum Bewilligungsvolumen der Deutschen Forschungsgemeinschaft erfolgen auf der Grundlage einer vierstufigen Fachsystematik, die die DFG zur Einteilung ihres Antragsaufkommens einsetzt. Quantitativ den größten Anteil am Fördervolumen der DFG nehmen Förderprogramme der sog. „Allgemeinen Forschungsförderung“ ein. Etwa zwei Drittel aller DFG-Mittel werden über die dort angebotenen Verfahren und Beihilfearten - etwa das „Normalverfahren“ (das in erster Linie der Finanzierung von Einzelprojekten dient), das „Schwerpunktverfahren“ (ein Programm zur gemeinsamen, instituts- und ortsübergreifenden Bearbeitung von durch Wissenschaftler vorgeschlagenen Forschungsthemen) oder die verschiedenen Stipendienprogramme der DFG - abgewickelt.

In der Mehrzahl der Förderprogramme der Allgemeinen Forschungsförderung erfolgt die fachliche Zuordnung eines Antrags zu operativen Zwecken. Orientiert an der inhaltlichen Ausrichtung eines Projekts wird ein Antrag einem von derzeit 189 Fächern zugewiesen. Innerhalb der Geschäftsstelle sind so genannte „Programmdirektoren“ für die Betreuung jeweils einer Gruppe verwandter Fächer zuständig. Mit der Zuordnung wird aber nicht nur die geschäftsstelleninterne Bearbeitung eines Antrags reguliert, auch der Begutachtungsprozeß ist damit in Teilen festgelegt. Gemäß ihrer Satzung beurteilen zunächst zwei Fachgutachter unabhängig voneinander einen Antrag. Diese Fachgutachter können in sehr unterschiedlichen Fächern tätig sein. Ein sogenannter „Fachausschussvorsitzender“, der mehreren Fächern (in jedem Fall aber dem Fach, dem ein Antrag zugeordnet wurde) vorsteht, formuliert auf der Grundlage dieser Gutachten eine abschließende Empfehlung. Diese wird durch regelmäßig tagende Gremien, denen sowohl Wissenschaftler (in der Mehrheit) als auch Vertreter des Bundes und der Länder angehören, zur Entscheidung gebracht.

Insgesamt hat die DFG genau 37 Fachausschüsse eingerichtet. Fachgutachter werden im 4-Jahres-Turnus von der Wissenschaftsgemeinschaft gewählt, Fachausschussvorsitzende und deren Vertreter wiederum werden von der Gruppe der zu einem Fachausschuß zusammengeschlossenen Fachgutachter bestimmt. In der aktuellen Wahlperiode (2000-2003) sind

insgesamt 650 Fachgutachter aktiv. In ihrer Arbeit werden sie von mit einem Thema besonders vertrauten, von Mitarbeitern der Geschäftsstelle ausgewählten „Sondergutachtern“ unterstützt⁵.

In den Programmen „Sonderforschungsbereiche“ und „Graduiertenkollegs“ erfolgt die fachliche Klassifikation einer Fördermaßnahme dagegen nur zu statistischen und dokumentarischen Zwecken. Die Wissenschaftler, die an diesen Programmen partizipieren, aber auch die Gutachterpanels, die die entsprechenden Einrichtungs- und Fortsetzungsanträge beurteilen, setzen sich - Ausdruck der vor allem diese Programme charakterisierenden Interdisziplinarität - aus Angehörigen eines oft sehr breiten Fächerspektrums zusammen. Im speziellen Fall von Sonderforschungsbereichen findet dies insofern Berücksichtigung, als nicht nur die Fördermaßnahme insgesamt - also der konkrete Sonderforschungsbereich - fachlich klassifiziert wird, sondern auch jedes einzelne Teilprojekt. Bewilligte Mittel für einen Sonderforschungsbereich können sich so auf mehrere Fachgebiete (und auch: auf mehrere beteiligte Einrichtungen) verteilen.

Die Klassifikation, die zur Abbildung der fachlichen Schwerpunkte dieser Programme eingesetzt wird, unterscheidet insgesamt 16 Fachgebiete. Diese unterteilen sich auf der einen Seite in die von oben bekannten 37 Fachausschüsse beziehungsweise 189 Fächer, auf der anderen Seite lassen sie sich zu insgesamt vier Wissenschaftsbereichen zusammenfassen. DFG-Fächer, Fachausschüsse, Fachgebiete und Wissenschaftsbereiche bilden gemeinsam die vier Stufen der Fachsystematik der DFG. Tabelle 2 gibt aus Gründen der Übersichtlichkeit allein die obersten drei Stufen dieser Systematik wieder.

⁵ Sondergutachter haben sich in den letzten Jahren zu einer wesentlichen Stütze des Begutachtungsverfahrens entwickelt. Aktuellen Statistiken zufolge, die zum Teil auf Schätzungen beruhen, werden pro Jahr zwischen 6.500 und 8.000 Sondergutachter mit der Vorbegutachtung von Anträgen betraut. Sondergutachter geben schriftliche Förderempfehlungen ab (die vor allem bei Einzelprojekten und Stipendien üblich sind), sind in großer Zahl aber auch an sog. „Panels“ beteiligt, in denen im Gruppengespräch über die Förderwürdigkeit von Maßnahmen diskutiert wird, die im Kontext koordinierter Programme (Forscherguppen, Schwerpunktprogramme, Sonderforschungsbereiche und Graduiertenkollegs) zur Entscheidung anstehen. 1999 wurden etwa 45 Prozent aller schriftlichen Förderempfehlungen von Sondergutachtern formuliert.

Fachausschuss		Fachgebiet	Wissenschaftsbereich
103	Rechtswissenschaft	Gesellschaftswissenschaften	Geistes- und Sozialwissenschaften
115	Geographie		
118	Wirtschaftswissenschaften		
119	Sozialwissenschaften		
107	Altertumswissenschaft	Geschichts- und Kunstwissenschaften	
108	Kulturen des Orients		
111	Geschichte		
112	Kunstwissenschaften		
113	Völkerkunde		
114	Geschichte der Naturwissenschaften		
109	Sprachwissenschaften A	Sprach- und Literaturwissenschaften	
110	Sprachwissenschaften B		
101	Ev. Theologie	Psychologie, Pädagogik, Philosophie, Theologie	
102	Kath. Theologie		
116	Philosophie		
117	Pädagogik		
120	Psychologie		
201	Theoretische Medizin	Medizin	Biologie und Medizin
202	Praktische Medizin		
205	Veterinärmedizin	Veterinärmedizin	
203	Biologie	Biologie	
207	Biologische Chemie und Physik		
204	Landwirtschaft und Gartenbau	Agrar- und Forstwissenschaften	
206	Forst- und Holzwissenschaften		
304	Mathematik	Mathematik	Naturwissenschaften
303	Physik	Physik	
302	Chemie	Chemie	
301	Wissenschaft der festen Erde	Geowissenschaften	
306	Hydrologie und Wasserwirtschaft		
401	Allg. Ingenieurwissenschaften	Allgemeine Ingenieurwissenschaften / Maschinenwesen	
408	Maschinenbau und Produktionstechnik		
409	Maschinenbau und Verfahrenstechnik		
402	Architektur, Städtebau und Landesplanung	Architektur, Städtebau, Bauingenieurwesen	
403	Bauingenieurwesen		
404	Bergbau und Hüttenwesen	Bergbau und Hüttenwesen	
406	Elektrotechnik	Elektrotechnik, Informatik	
407	Informatik		

Tabelle 2: DFG-Systematik der Fachausschüsse, Fachgebiete und Wissenschaftsbereiche

Mit Hilfe dieser Systematik läßt sich praktisch jeder an die DFG gerichtete Antrag zuordnen. Ausnahmen bilden vor allem Maßnahmen im Bereich der Bibliotheksförderung sowie einzelne Hilfseinrichtungen mit fachübergreifenden Zielsetzungen.

4.4 Absolutes und relatives DFG-Drittmittelvolumen im Vergleich

Das Hauptgewicht des „DFG-Rankings“ liegt auf der Offenlegung des absoluten DFG-Bewilligungsvolumens je Hochschule und Fachgebiet. Die in entsprechender Form aufgeschlüsselten Daten geben Aufschluß über lokale Schwerpunktsetzungen DFG-finanzierter Forschung und erlauben so, das fachliche „DFG-Profil“ einer Hochschule über Lehrstuhl- und Fakultätsgrenzen hinweg zu identifizieren. Als besonders intensiv erweist sich so etwa die gesellschaftswissenschaftliche Forschung in Bonn, Frankfurt und Berlin (HU), während die Zentren DFG-finanzierter Sprach- und Literaturwissenschaft eher in München (U), Konstanz und Tübingen verortet sind. In der Medizin gehen die höchsten Bewilligungsvolumina an die Universitäten in Berlin (HU), München (U) und Würzburg, die Rangfolge biologisch forschender Standorte wird von München (U), Würzburg und Heidelberg angeführt. In der Mathematik liegen Heidelberg, Bonn und Bielefeld mit weitem Abstand vorn, in der Chemie sind es München (TU), Marburg und Mainz. Die Allgemeinen Ingenieurwissenschaften, ein Fachgebiet, das einen sehr weiten Fächerkanon zusammenfasst, weisen als ihre Zentren die Universitäten in Stuttgart, Aachen und Hannover aus, das im Verhältnis hierzu relativ kleine Fachgebiet Elektrotechnik/Informatik führt an der Spitze die Universitäten in München (TU), Karlsruhe, Dresden und Paderborn.

Weil jede Hochschule eigene Akzente setzt, finden sich je Fachgebiet ganz eigene „Spitzengruppen“ beziehungsweise - auf der anderen Seite - je eigene Gruppen von Hochschulen, die „auf den Plätzen“ folgen. Daß dabei keine Aussagen über Institute und Fachbereiche sondern über DFG-Fachgebiete getroffen werden, markiert eine eigene Qualität des „DFG-Rankings“: Fakultäten und Fachbereiche binden oft eine Reihe von Lehr- und Forschungsgebieten ein, die in ihrer speziellen Zusammensetzung über lokale Grenzen hinaus kaum eine Entsprechung finden. Firmiert hier die Psychologie unter den Geisteswissenschaften, ist sie dort an einem gesellschaftswissenschaftlichen, an anderer Stelle wiederum an einem medizinischen Fachbereich angesiedelt. Erfolgt hier medizinische Forschung zuvorderst im Kontext universitärer Kliniken, ist sie dort vor allem den an einem biologischen Fachbereich verorteten Wissenschaftlern zu verdanken. Das „DFG-Ranking“ ignoriert solchermaßen an lokalen Traditionen orientierte, oft willkürlich anmutende „Grenzziehungen“ und hat bei einer Zuordnung allein die in Projektanträgen formulierten Forschungsziele im Auge. Ob ein Projekt in der Informatik an einem Lehrstuhl für Informatik, für Medizin (etwa: Medizinische Informatik), für Betriebswirtschaft (Wirtschaftsinformatik) oder gar für Rechtswissenschaften (Rechtswissenschaften) durchgeführt wird, ist dabei ohne Belang. Das DFG-Ranking gibt so auch keinen Hinweis auf die Leistungsfähigkeit bestimmter Organisationseinheiten vor Ort. Es informiert vielmehr über fachliche Schwerpunktsetzungen, die instituts- und fakultätsübergreifend gesetzt werden.

Die Relevanz der auf absolute Volumina fokussierten Berichterstattung zeigt sich am ehesten dort, wo über einzelne Einrichtungen hinaus Aussagen zu regionalen Schwerpunktsetzungen getroffen werden. So dokumentiert das aktuelle DFG-Ranking beispielsweise auch, in welchem Umfang Mittel in nach vier Wissenschaftsbereichen und 16 Fachgebieten differenzierender Form in bestimmte „Forschungsregionen“ Deutschlands geflossen sind. Berichtseinheiten

bilden PLZ-Gebiete, operationalisiert werden diese über die ersten beiden Stellen der Postleitzahl eines Ortes, an die eine Bewilligung ausgesprochen wurde⁶.

Die kartographischen Analysen weisen je Fachgebiet markante Schwerpunktsetzungen aus. Sie geben so Anhaltspunkte über die Intensität, mit der in diesen Regionen DFG-finanzierte Forschung betrieben wird. Als eine Hochburg für geistes- und sozialwissenschaftliche Forschung erweist sich etwa Berlin, wo nicht nur drei große Universitäten sondern eine ganze Reihe außeruniversitärer Forschungseinrichtungen - etwa das „Deutsche Institut für Wirtschaftsforschung (DIW)“ (das unter anderem das sozialwissenschaftliche Großprojekt „Sozio-Ökonomisches Panel“ betreut), drei „Geisteswissenschaftliche Zentren“, das „Wissenschaftszentrum Berlin für Sozialforschung (WZB)“, das „Deutsche Archäologische Institut“ sowie verschiedene Einrichtungen der Max-Planck-Gesellschaft - zu einem für dieses Gebiet insgesamt stimulierenden Forschungsklima beitragen. Ähnliche „Ballungsräume“ lassen sich für die drei anderen Wissenschaftsbereiche, beziehungsweise, in detaillierterer Form, für jedes der 16 Fachgebiete der DFG-Systematik, identifizieren.

Aus der räumlichen Nähe fachlich verwandter Forschungseinrichtungen ergeben sich gute Rahmenbedingungen für örtliche Schwerpunktsetzungen. Ist eine Region als „führend“ auf einem bestimmten Fachgebiet anerkannt, wirkt sie wie ein Magnet auf Wissenschaftler, die auf diesem Gebiet reüssieren wollen. Der Begriff des „Forschungszentrums“ findet vor allem hier, in der auf Regionen zielenden Sichtweise, seinen Bezugspunkt.

Ein Manko des Organisationseinheiten übergreifenden Charakters des „DFG-Rankings“ liegt in der Tatsache begründet, daß es kaum möglich ist, DFG-Bewilligungen für bestimmte Fachgebiete zur Zahl der auf diesem Fachgebiet insgesamt tätigen Wissenschaftler in Beziehung zu setzen. Zwar liegen durch das Statistische Bundesamt für jede einzelne Hochschule sehr ausdifferenzierte Statistiken zur Zahl der Wissenschaftler je sog. „Lehr- und Forschungsbereich“ vor, die sich mit hinreichender Genauigkeit zu DFG-Fachgebieten in Beziehung setzen lassen⁷; da diese Systematik organisatorische Ordnungsprinzipien wiedergibt, für die DFG aber - wie an Beispielen illustriert - allein inhaltliche Kriterien für die Zuordnung einer Fördermaßnahme zu einem Fachgebiet maßgeblich sind, lassen sich Wissenschaftlerzahlen je Lehr- und Forschungsbereich nicht 1:1 auf DFG-Fachgebiete abbilden.

Die im „DFG-Ranking“ vorgenommenen Relativierungen beschränken sich aus diesem Grund auf die Gegenüberstellung von DFG-Bewilligungen je Wissenschaftsgebiet und den in Fächern dieses Gebiets an einer Hochschule tätigen Personen. Auf dieser hoch aggregierenden Ebene sind relativierende Aussagen hinlänglich belastbar. Die Bezugnahme auf die Zahl der an einer Hochschule in einem bestimmten Wissenschaftsbereich aktiven Wissenschaftler erfolgt, wie bereits in Tabelle 1 dargestellt, für zwei Personengruppen: 1. Wissenschaftler insgesamt und 2. Professoren. Diese Unterscheidung trägt fachgebietspezifischen Kulturen Rechnung.

⁶ Kartographische Darstellungen für vier Wissenschaftsbereiche finden sich in der gedruckten Version des Berichts. Die Online-Präsentation zu diesem Bericht stellt darüber hinaus Detailkarten für insgesamt 16 Fachgebiete zur Verfügung (vgl. <http://www.dfg.de/berichtswesen/ranking.html>).

⁷ Vgl. Tabelle 26 aus: DFG 2000: DFG-Bewilligungen an Hochschulen und außeruniversitäre Forschungseinrichtungen: Seite 111-112.

	Professoren	Wiss. Personal insgesamt	Anteil Professoren in %
Geistes- und Sozialwissenschaften	8.238	30.688	26,8
Biologie / Medizin	4.827	47.644	10,1
Naturwissenschaften	3.909	23.185	16,9
Ingenieurwissenschaften	3.170	22.477	14,1
Insgesamt	20.144	123.994	16,2

Quelle: Statistisches Bundesamt, Hauptberuflich tätiges wissenschaftliches und künstlerisches Personal an Hochschulen (Stand: 1998). Sonderauswertung im Auftrag von DFG und HRK (Basis: 72 Universitäten, die zwischen 1996 und 1998 mehr als 3 Millionen DM DFG-Bewilligungen erhalten haben).

Tabelle 3: Professorenanteil am wissenschaftlichen Personal je Wissenschaftsbereich, Stand: 1998 (DFG 2000: 32).

Tabelle 3 weist für die vier Wissenschaftsbereiche der DFG-Systematik deutliche Unterschiede im Anteil aus, den Professoren an Wissenschaftlern insgesamt einnehmen. Allgemein liegt der Professorenanteil an den betrachteten Universitäten bei 16 Prozent. Am niedrigsten fällt er in den biologisch-medizinischen Fachgebieten aus, dort zählt nur jeder zehnte der an einer Hochschule tätigen Wissenschaftler zu dieser Statusgruppe. In den geistes- und sozialwissenschaftlichen Fächern ist dagegen mehr als jeder vierte Wissenschaftler der Gruppe der Professoren zuzurechnen.

In gewissem Umfang spiegeln sich diese Unterschiede auch in der Zusammensetzung der Antragstellerschaft der DFG wider. Analysen auf der Basis von Sachbeihilfeanträgen im Normal- und Schwerpunktverfahren zeigen, dass etwa im Fachgebiet Medizin knapp 60 Prozent aller Antragsteller keinen Professorenstatus aufweisen. In diesem Gebiet tritt also überwiegend der wissenschaftliche Nachwuchs mit Anträgen an die DFG heran. Im ingenieurwissenschaftlichen Bereich stellen dagegen in erster Linie Professoren Anträge. Besonders hoch ist deren Anteil etwa in den Fachgebieten „Elektrotechnik, Informatik“ und „Architektur, Städtebau, Bauing.-wesen“, wo sie für 90 Prozent aller Anträge verantwortlich zeichnen (DFG 2000: 33).

Welche Unterschiede ergeben sich, wenn man je Wissenschaftsbereich die hier diskutierten Rangreihen gegenüberstellt? Rangieren große Universitäten, die absolut ein hohes Bewilligungsvolumen einwerben, in der auf die Professorenzahl relativierten Sicht „auf den Plätzen“? Zeigen sich weiterhin signifikante Unterschiede, je nachdem, ob Bewilligungsvolumina zur Zahl der Professoren oder zur Zahl der Wissenschaftler insgesamt ins Verhältnis gesetzt werden?

Geistes- und Sozialwissenschaften			Biologie/Medizin			Naturwissenschaften			Ingenieurwissenschaften		
Universität	P	W	Universität	P	W	Universität	P	W	Universität	P	W
München(U)			München (U)			Hamburg (U)			Aachen		
Berlin (HU)			Würzburg			Bonn			Stuttgart		
Berlin (FU)			Heidelberg			Mainz			Karlsruhe		
Tübingen			Berlin (HU)			München (TU)			München (TU)		
Frankfurt			Tübingen			Heidelberg			Hannover (U)		
Bonn			Göttingen			Göttingen			Berlin (TU)		
Bielefeld			Freiburg			Berlin (TU)			Erlangen-Nürnberg		
Köln			Marburg			Berlin (FU)			Darmstadt		
Heidelberg			Düsseldorf			Karlsruhe			Braunschweig		
Konstanz			Berlin (FU)			Tübingen			Dresden		

P=Bewilligungsvolumen im Verhältnis zur Zahl der Professoren

W=Bewilligungsvolumen im Verhältnis zur Zahl der Wissenschaftler insgesamt

	Platz 1-10 der Vergleichsreihe
	Platz 11-20 der Vergleichsreihe
	Platz 21 und folgende der Vergleichsreihe

Tabelle 4: Absolutes und relatives DFG-Bewilligungsvolumen der zehn größten Bewilligungsempfänger je Wissenschaftsbereich im Vergleich

Tabelle 4 zeigt, dass die Antwort auf diese Fragen je nach Wissenschaftsbereich unterschiedlich ausfällt. Aufgeführt sind dort zunächst jene zehn Universitäten, die absolut das höchste DFG-Bewilligungsvolumen je Wissenschaftsbereich eingeworben haben. In den mit „P“ überschriebenen Spalten informieren Muster codes über den Rangbereich, den eine Universität besetzt, wenn man deren absolutes Bewilligungsvolumen zur Zahl der in einem Wissenschaftsbereich tätigen Professoren in Beziehung setzt. Die Spalte „W“ informiert dementsprechend über den Rangbereich in einer zur Zahl der insgesamt an einer Universität tätigen Wissenschaftler relativierenden Rangfolge. Weiß markierte Felder symbolisieren, dass eine Universität auch in der entsprechenden Referenzrangreihe einen der 10 ersten Plätze belegt, gestreifte Markierungen werden für Hochschulen verwendet, die dort auf den Plätzen 11 bis 20 rangieren, eine schwarze Markierung wird schließlich zugewiesen, wenn ein Rang auf Platz 21 und folgende dokumentiert ist.

Für die Geistes- und Sozialwissenschaften und auch für die Naturwissenschaften ergibt die Gegenüberstellung ein relativ „harmonisches“ Bild: Ob DFG-Bewilligungen an eine Universität absolut oder in Relation zu den dort tätigen Professoren oder Wissenschaftlern insgesamt in Beziehung gesetzt werden, führt mit Blick auf die zehn führenden Universitäten nur im Ausnahmefall zu einschneidenden Unterschieden. In den Geistes- und Sozialwissenschaften finden sich sieben von zehn die „absolute“ Rangreihe anführenden Universitäten auch in den zur Zahl der in diesen Wissenschaftsbereichen tätigen Wissenschaftler relativierenden

Übersichten unter den „Top Ten“. In den Naturwissenschaften ist das Ergebnis ähnlich eindeutig (Übereinstimmung in acht (Bezugsgröße Professoren), beziehungsweise sieben (Wissenschaftler insgesamt) Fällen). Nur jeweils eine Universität fällt in den auf Wissenschaftler insgesamt bezogenen Rangreihen etwas weiter ab, in den Geistes- und Sozialwissenschaften handelt es sich um die Universität zu Köln, in den Naturwissenschaften ist es die TU München.

Ein etwas „bunteres“ Bild ergibt sich für die Ingenieurwissenschaften. Dort verbleiben in den Referenzrangreihen nur noch sechs (Bewilligungen je Professor) beziehungsweise drei (Bewilligungen je Wissenschaftler insgesamt) Universitäten unter den zehn Erstplazierten, zwei beziehungsweise drei Universitäten fallen gar auf weit vom Spitzenfeld entfernte Plätze (Rang 21 und folgende) ab.

Regelrecht divergent sind die Ergebnisse schließlich für den Wissenschaftsbereich Biologie/Medizin. Dort lassen die verschiedenen Rangreihen kaum noch Ähnlichkeiten erkennen. Allein die Universitäten in Marburg und Würzburg sind sowohl absolut als auch relativ zur Professorenzahl in der Spitzengruppe verortet. Zwischen der absoluten Rangreihe und der zur Zahl der Wissenschaftler insgesamt relativierenden Folge ergibt sich mit Blick auf die Spitzengruppe keinerlei Übereinstimmung. Vor allem für die Lebenswissenschaften ist es damit von großer Bedeutung, ob absolute oder relative Zahlen zur Beurteilung der DFG-Aktivität einer Hochschule herangezogen werden.

5. Ausblick

Mit dem so genannten „DFG-Ranking“ hat die Geschäftsstelle der Deutschen Forschungsgemeinschaft einen Service etabliert, der auch in Zukunft einen festen Bestandteil ihres Informationsangebots bilden wird. Die große Aufmerksamkeit, die diese Übersichten bisher erfahren haben, trägt dazu bei, daß die Geschäftsstelle der DFG das in ihren Förderdatenbanken archivierte „Wissen“ über Forschung in Deutschland insgesamt neu bewertet. Diese Datenbanken werden heute nicht länger nur zur Unterstützung interner Geschäftsabläufe genutzt, sondern bilden die Grundlage für ein zunehmend breiter werdende Dienstleistungsangebot. Eine aktuelle Neuerung bietet etwa ein Informationssystem, das via Internet die Stichwortrecherche in Abstracts zu DFG-geförderten Projekten ermöglicht⁸. Weitergehend als dies ein „Ranking“ je erlauben würde, werden sich mit dieser Datenbank mittelfristig in detaillierter Form Aussagen zu DFG-finanzierten Forschungsschwerpunkten treffen lassen. Lässt ein Ranking allenfalls quantifizierende Aussagen über Forschungsaktivitäten in bestimmten Fachgebieten und an bestimmten Einrichtungen zu, können anhand dieser Quelle auch qualifizierende Informationen zu Forschungsaktivitäten in individuell eng definierten Themenfeldern gewonnen werden: Aussagen sind dann beispielsweise nicht mehr allein zu den Standorten „psychologischer“ Forschung möglich, sondern zu den Personen und Personengruppen, die auf bestimmten Spezialgebieten, etwa der Schizophrenie- oder der Zwillingforschung aktiv sind. Nicht mehr die Ingenieurwissenschaften in ihrer Gesamtheit, sondern im speziellen die Materialforschung (mit ihren stark

⁸ Vgl. <http://www.dfg.de/gepris> ;

interdisziplinären Bezügen zur biologischen und vor allem chemischen Forschung) gerät in den Blick ebenso wie beispielsweise die Ernährungsforschung, die in ähnlicher Weise von Angehörigen unterschiedlichster Disziplinen betrieben wird.

Weiterentwicklungen sind aber auch für das „DFG-Ranking“ selbst vorgesehen. Auf der technischen Seite werden derzeit Bedingungen geschaffen, die es in Zukunft erlauben werden, die Erstellung entsprechender Übersichten sowohl mit geringerem Aufwand als auch mit erhöhter Tiefe zu bewerkstelligen. Die Reduzierung des Aufwandes wird sich in erster Linie aus einer ab 2002 wirksam werdenden, alle Förderprogramme der DFG umfassenden, einheitlichen Datenhaltung ableiten (derzeit werden Daten noch in einer Vielzahl von Quellen (mit je eigenen Datenformaten) archiviert). „Erhöhte Tiefe“ heißt, dass es nach dieser sehr aufwendigen Umstellungen beispielsweise auch möglich sein soll, Informationen zum Institut eines Antragstellers in einer für statistische Analysen geeigneten Form in der Datenbank abzubilden - was es dann auch erlauben würde, Daten zu Bewilligungen je Institut und/oder Fachbereich/Fakultät bereitzustellen. Schließlich soll es dann auch möglich sein, nicht nur über Bewilligungen, sondern über tatsächlich in einem Haushaltsjahr abgerufene Mittel zu berichten.

Die Erfüllung der beiden letztgenannten Punkte bildet eine zentrale Voraussetzung, um Daten der DFG zu Daten anderer Drittmittelgeber in Beziehung zu setzen. Die DFG beteiligt sich derzeit aktiv an Diskussionen, die in diesem Punkt eine engere Zusammenarbeit der mit Forschungsförderung befassten Stellen in Deutschland zum Ziel haben. Angestrebt wird die Etablierung von Strukturen, die - etwa in Form einer zentralen Evaluationsagentur - zu einer verbesserten Transparenz und Übersichtlichkeit beziehungsweise zur Bereitstellung von für Forschungsplanung und objektive Leistungsbewertung gleichermaßen wichtigen Kennzahlen führen. Die DFG, dies ist nicht zuletzt in der Aufmerksamkeit begründet, die dem „DFG-Ranking“ in der wissenschaftlichen und außerwissenschaftlichen Öffentlichkeit zuteil wird, wird in dieser Diskussion eine wichtige Rolle einnehmen - gemeinsam mit anderen Wissenschaftsorganisationen, die das Interesse an einer das Profil des Forschungsstandorts Deutschland schärfenden Entwicklung eint.

Literatur:

Wissenschaftsrat: *Drittmittel und Grundmittel der Hochschulen 1993 - 1998*, Köln 2000.

Deutsche Forschungsgemeinschaft: *DFG-Bewilligungen an Hochschulen und außeruniversitären Forschungseinrichtungen - 1996 bis 1998*, Bonn 2000.

Aussagefähigkeit von Patentanalysen zur Bewertung von Hochschulen

Ulrich Schmoch

Einleitung

Noch vor wenigen Jahren war es selbstverständlich, die Leistung von Hochschulen nach der Qualität von Lehre und Grundlagenforschung zu beurteilen. In letzter Zeit wird aber auch verstärkt nach der Bedeutung der wissenschaftlichen Forschung für die Gesellschaft gefragt, weshalb die Aktivitäten der Hochschulen im Bereich der angewandten Forschung größere Beachtung finden. Dieses gewachsene Interesse für die unmittelbare Nützlichkeit der Hochschulforschung ist auf die steigenden Kosten moderner Forschung bei gleichzeitiger Mittelverknappung der öffentlichen Hand zurückzuführen. Außerdem ist eine wachsende Relevanz so genannter wissensintensiver Technologien zu beobachten, deren Entwicklung maßgeblich auf den Ergebnissen der Hochschulforschung beruht (vgl. Stankiewicz 1986: 5 ff; Grupp/ Schmoch 1992). Schließlich spielt die Frage der Anwendungsorientierung auch bei der Hochschulwahl von Studienanfängern eine Rolle, die in letzter Zeit größeren Wert auf den Aspekt der Berufsausbildung gegenüber dem der wissenschaftlichen Qualifizierung legen (Bierhals/Schmoch 2000: 82). Ein weiterer Trend der letzten Jahre ist das Bemühen, die Leistungen der Hochschulen und ihrer Fachbereiche und Lehrstühle durch quantitative Indikatoren abzubilden, um so eine rationale Grundlage für die Strategieentwicklung und die Mittelverteilung bereit stellen zu können (Fuhrmann et al. 2000).

Zur Messung der Ausrichtung auf die angewandte Forschung sind Patente aus Hochschulen als möglicher Indikator vorgeschlagen worden, bei denen eine steigende Zahl von Anmeldungen zu beobachten ist. Im folgenden wird näher analysiert, ob Patentanmeldungen aus Hochschulen tatsächlich ein geeigneter Indikator zur Bewertung der Anwendungsorientierung sind. Dabei handelt es sich nicht um den Versuch, die Leistung der Hochschulen insgesamt allein auf der Basis von Patentanmeldungen zu bewerten. Vielmehr geht es lediglich um die Bewertung eines Teilbereichs der Hochschulforschung (Pavitt 1998), der allerdings von steigender Relevanz ist. In der Untersuchung wird auch nicht der Frage nachgegangen, ob ein zu starkes Absichern von Ergebnissen der Hochschulforschung über gewerbliche Schutzrechte volkswirtschaftlich sinnvoll ist (Mazzoleni/Nelson 1998; Ganz-Brown 1999).

Die steigende Zahl von Patentanmeldungen aus Hochschulen darf nicht nur als Indikator für eine stärkere Orientierung der Hochschulforschung auf unmittelbar anwendungsrelevante Ergebnisse interpretiert werden. Vielmehr reflektiert sie auch ein gewachsenes Bewusstsein für die Möglichkeit der privaten Aneignung von Forschungsergebnissen und damit einen Wandel von Institutionen der Wissenschaft. So werden die von Merton (<1942>1985) als typisch herausgestellten Institutionen des Kommunalismus und der Uneigennützigkeit zumindest partiell durch Privatisierung und Eigeninteresse verdrängt (Felt et al. 1995: 63). In der Sichtweise von Spinner (1994) ist die Trennung von Erkenntnis und Eigentum eine typische Institution der "alten Wissensordnung", die in der "neuen Wissensordnung" durch deren Konjunktion ersetzt wird. In jeden Fall spiegeln Patentanmeldungen aus Hochschulen einen tiefer liegenden Wandel, der im

Kontext der aktuellen Diskussion über eine Reform der Universitäten zu sehen ist (Stifterverband 2000).

1. Entwicklung von Patentanmeldungen aus Hochschulen

Im Patentgesetz sind als Voraussetzungen für die Patentierung einer Erfindung die weltweite Neuheit, eine erfinderische Tätigkeit (Erfindungshöhe) sowie die gewerbliche Anwendbarkeit festgelegt. Das Kriterium der gewerblichen Anwendbarkeit zeigt, dass es nicht nur um die Zertifizierung der Neuheit einer Idee geht, sondern jede Patentanmeldung auf eine Einführung der entsprechenden Technologie in den Markt zielt (Schmoch et al. 1988: 29; Schmoch 1990: 15 ff). Von daher ist es konsequent, wenn es im Patentgesetz heißt:

"Als Erfindungen ... werden insbesondere nicht angesehen ... Entdeckungen sowie wissenschaftliche Theorien und mathematische Methoden" (PatG 1981 §1 Abs. 2 zitiert nach DPMA 2000).

Eine klare Ausrichtung auf eine kommerzielle Verwertung von Erfindungen ergibt sich auch aus den Kosten des Patentverfahrens, bei dem amtliche Gebühren, Patentanwaltskosten und im Falle von Auslandsanmeldungen auch Übersetzungskosten anfallen. So beläuft sich eine deutsche Anmeldung bis zur Erteilung auf Gesamtkosten von etwa 10.000 DM; eine Anmeldung am US-amerikanischen Patentamt (USPTO) kostet etwa 15.000 DM, eine Anmeldung am Europäischen Patentamt (einschl. der Überleitung an nationale Ämter) 40.000 DM und mehr.

Aus den oben skizzierten Rahmenbedingungen lassen sich verschiedene Voraussetzungen ableiten, die für Patentanmeldungen aus Hochschulen maßgeblich sind. Zunächst einmal muss ein praxisrelevantes, technisches Ergebnis der Hochschulforschung vorliegen, das sich überhaupt für die Anmeldung eines Patentes eignet. Als zweites muss die Bereitschaft des Hochschullehrers (Forschers) vorliegen, ein Patent anzumelden um damit den Weg für eine private Aneignung des Wissens zu öffnen. Als dritte Voraussetzung muss es eine Möglichkeit zur Finanzierung der Verfahrenskosten geben. Dabei muss der Hochschullehrer entweder bereit sein, diese Kosten privat zu tragen, oder er muss ein Unternehmen finden, das diese Last übernimmt. Dieses führt zur vierten Voraussetzung für Patentanmeldungen aus Hochschulen. Der Hochschullehrer muss ein oder mehrere Unternehmen kennen, die an einer Verwertung interessiert sind, oder er muss zumindest gute Aussichten auf potentielle Verwerter haben. Denn die Hochschule selbst verfügt nicht über die Ressourcen, die für die Einführung einer Erfindung am Markt erforderlich sind.

Trotz dieser Vielzahl von Restriktionen ist die Bereitschaft von Hochschullehrern zur Durchführung von Patentanmeldungen überraschend hoch. Bei einer Befragung von Hochschullehrern zeigte sich, dass immerhin 43 Prozent im Zeitraum von 1990 bis 1997 patentfähige Forschungsergebnisse erzielt und diese auch zum Patent angemeldet haben. Weitere 24 Prozent verfügten zwar über patentfähige Forschungsergebnisse, haben aber kein Patent angemeldet. Die Restgruppe mit einem Anteil von 33 Prozent erzielte keine patentfähige Forschungsergebnisse und hat damit auch keine Patente angemeldet (Krause/Schroll 1999: 23). Bei der Interpretation dieser Ergebnisse ist zu berücksichtigen, dass die Befragung zu jeweils einem Drittel auf Hochschullehrer in den Ingenieurwissenschaften, Naturwissenschaften und Biowissenschaften zielte, also die Gruppe betraf, bei der überhaupt Patentanmeldungen denkbar sind.

Angesichts dieses Befundes entspricht es den Erwartungen, dass die Zahl der Patentanmeldungen aus deutschen Hochschulen seit Beginn der 70er Jahre kontinuierlich angestiegen ist (Abbildung 1). Die steigende Zahl von Patentanmeldungen aus Hochschulen impliziert auch eine steigende Bedeutung der Hochschulen für das deutsche Innovationssystem. Der Anteil von Hochschulanmeldungen an allen Anmeldungen deutscher Herkunft ist seit Anfang der 70er Jahre von 1,5 Prozent auf aktuell 4,5 Prozent gestiegen. Es ist allerdings denkbar, dass diese Steigerung nur bedingt eine stärkere Anwendungsorientierung der Hochschulen widerspiegelt, sondern in erster Linie auf ein Wachstum der Hochschulforschung zurückzuführen ist. Werden die Zahlen in Relation zu den Forschungsaufwendungen an Hochschulen gesetzt, zeigt sich tatsächlich, dass es unterschiedliche Phasen gibt. Während in den 70er Jahren eine Steigerung der Zahl der Patente bezogen auf die Forschungsausgaben zu beobachten ist, steigt diese Quote in den 80er Jahren kaum an. Insbesondere seit 1994 ist aber wieder ein erheblicher Zuwachs festzustellen, so dass für die jüngste Zeit ein erhebliches Interesse der Hochschullehrer an Patentanmeldungen konstatiert werden kann.

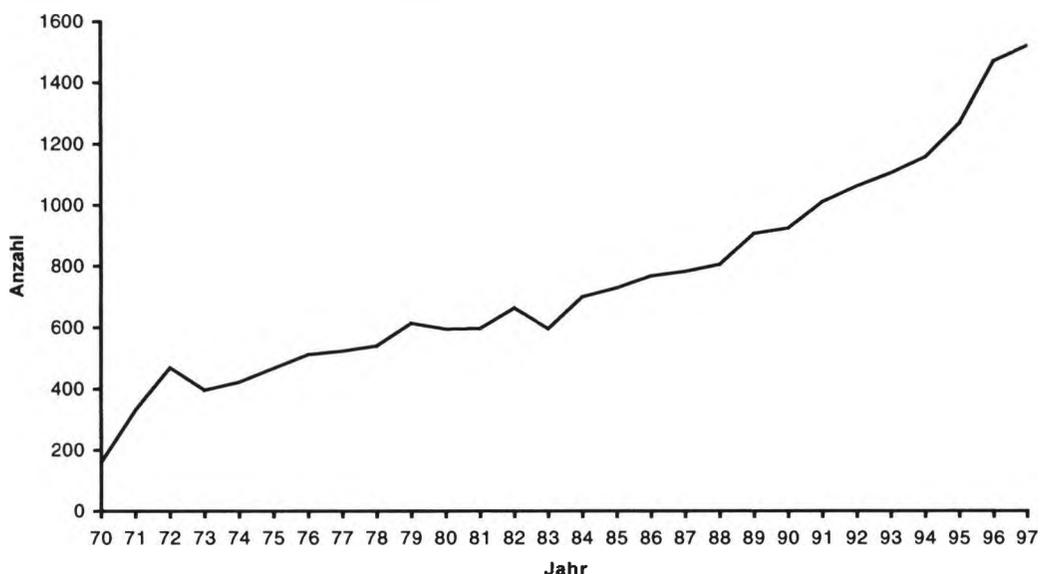


Abbildung 1:

Die Recherche nach Patentanmeldungen aus Hochschulen ist aus methodischer Sicht schwierig, weil nach der derzeitigen Rechtslage Hochschullehrer privat über ihre Erfindungen verfügen können und daher eine Hochschule nur in Ausnahmefällen als Patentanmelder in Erscheinung tritt (Schmoch 2000a; Becher et al. 1996: 68 ff). Die in Abbildung 1 dargestellten Recherche-Ergebnisse beruhen darauf, dass der Anmelder oder Erfinder eine Professorentitel hat, womit zumindest ein Großteil der Patentanmeldungen aus Hochschulen erfasst wird.¹ Nicht erfasst sind dagegen Anmeldungen aus Hochschulen, bei denen die Erfinder keinen Professorentitel tragen oder diesen nicht angegeben haben, so dass die Zahlen insgesamt unterschätzt sind. Es

¹ Zur Methodik vgl. insbesondere Becher et al. (1996: 23ff). In der Studie von Becher et al. war der Autor für die patentstatistischen Analysen verantwortlich. In dieser Stichprobe sind Anmeldungen ausgeschlossen, in der die Erfinder nicht an Hochschulen, sondern an anderen Forschungseinrichtungen oder in Unternehmen arbeiten.

dürfte aber der größte Teil der relevanten Stichprobe erfasst und der ermittelte Aufwärtstrend statistisch signifikant sein.

Eine manuelle Auswertung der Patentanmeldungen aus Hochschulen für das Jahr 1993 hat gezeigt, dass nur 35 Prozent von den Hochschullehrern privat angemeldet werden, 6 Prozent von Hochschulen und 5 Prozent von mit der Hochschule kooperierenden Forschungseinrichtung. Das bedeutet, dass von vornherein 54 Prozent aller Anmeldungen aus Hochschulen von Unternehmen angemeldet werden, was die oben beschriebene Bedeutung von interessierten Verwertungspartnern eindrucksvoll illustriert (Becher et al. 1996: 28 f).

2. Disaggregierte Auswertungen von Hochschul-Patentanmeldungen

Die ermittelte Stichprobe von Patentanmeldungen aus Hochschulen erweist sich als hinreichend groß, um weitere disaggregierte Analysen durchzuführen. So eignet sich der Datensatz zur vertieften Untersuchung der Frage, in welchen Disziplinen der deutschen Hochschulforschung Ergebnisse erarbeitet werden, die sich für eine unmittelbare kommerzielle Verwertung eignen. Zu diesem Zweck wurde die Stichprobe für das Jahr 1993 genauer im Hinblick auf die fachliche Herkunft der Erfinder ausgewertet. Da sich in den Patentanmeldungen selbst keine entsprechenden Angaben finden, war eine manuelle Auswertung mit Hilfe von einschlägigen Hochschullehrer-Verzeichnissen erforderlich. Im Ergebnis zeigt sich eine Dominanz des Maschinenbaus und der Chemie, die zusammen fast die Hälfte aller Patentanmeldungen aus Hochschulen umfassen (Tabelle 1).² Auch aus der Medizin, Elektrotechnik und Physik kommen noch nennenswerte Beiträge zu Patentanmeldungen aus Hochschulen. Als problematisch bei der Zuordnung der Disziplinen erwiesen sich die Materialforschung, die teils eigenständig, teils mit Chemie oder Maschinenbau assoziiert ist, sowie die Pharmazie, die oftmals nicht eigenständig, sondern als Teil der Chemie ausgewiesen ist.

² Die größere Bedeutung des Maschinenbaus im Vergleich zur Chemie bei Krause/Schroll (1999: 17) ist auf die dort gewählte Stichprobenauswahl zurückzuführen. Die hier betrachtete Vollausswertung des Jahrgangs 1993 ist repräsentativer.

	<i>Absolut</i>	<i>Relativ (%)</i>
1. Maschinenbau	252	24
2. Chemie	237	23
3. Medizin	156	15
4. Elektrotechnik	101	10
5. Physik	84	8
6. Biologie	42	4
7. Landwirtschaft	39	4
8. Bauwesen	27	3
9. Materialforschung	23	2
10. Sonst. Ing.-wiss.	19	2
11. Pharmazie	17	2
12. Informatik	11	1
13. Geowissenschaft	4	0
14. Mathematik	3	0
15. Andere	18	2
Total	1033	100

Tabelle 1: Patentanmeldungen aus Hochschulen

Als wesentliches Ergebnis ist festzuhalten, dass in Deutschland die Hochschulen einen wesentlichen Beitrag zu Innovationen im Maschinenbau leisten. Dieses ist keineswegs selbstverständlich, da der Maschinenbau in der internationalen Diskussion als weniger wissensintensive Technologie gilt, bei der die Bedeutung der Hochschulforschung begrenzt ist (Schmoch 1997: 108). So stammt in den Vereinigten Staaten der größte Teil der Hochschulanmeldungen aus den Bereichen Chemie, Medizin, Mikroelektronik und Datenverarbeitung (Ganz-Brown 1999: 407). Die Tatsache einer großen Bedeutung der Interaktion zwischen Hochschulen und Unternehmen im Maschinenbau, wie sie sich in den Patentanmeldungen widerspiegelt, ist nicht von vornherein als Defizit des deutschen Innovationssystems zu interpretieren. Es gibt vielfältige Gründe, die solche Beziehungen sinnvoll machen und als Stärken der deutschen Strukturen gewertet werden können. Für die vorliegende Fragestellung ist jedoch wichtig, dass die Patentanmeldungen aus Hochschulen derartige Analysen auf der Makroebene ermöglichen und diese Besonderheit der deutschen Situation erfassen können. Insbesondere eröffnet sich über Patentanmeldun-

gen auch die Möglichkeit, die Veränderung der disziplinären Verteilung im Zeitverlauf zu beobachten.

Werden die Patentanmeldungen nach Hochschulen differenziert, lassen sich besonders patentaktive ermitteln. Tabelle 2 zeigt eine entsprechende Aufstellung für die Jahre 1985 und 1993, die auf dem oben beschriebenen Verfahren einer manuellen Auswertung der jeweiligen Stichproben beruht. Generell zeigt sich zwischen den beiden Vergleichsjahren eine aufsteigende Tendenz; im Falle der RWTH Aachen, der TU Berlin und der Universität Stuttgart sind die Zahlen dagegen rückläufig. Diese zeitlichen Veränderungen sind ohne eine Kenntnis der genaueren Umstände nur schwer interpretierbar. Die oben diskutierten Voraussetzungen für Patentanmeldungen aus Hochschulen legen es nahe, dass diese von einem Zusammentreffen verschiedener singulärer Ereignisse abhängen und damit ein in der Zeit erratischer Verlauf zu vermuten ist.

		Patentanmeldungen	
		1993	1985
1.	RWTH Aachen	61	87
2.	TU München	48	32
3.	Universität Karlsruhe	42	24
4.	TU Dresden	38	128*
5.	Universität Erlangen-Nürnberg	36	17
6.	Universität Göttingen	35	7
7.	FU Berlin	33	15
8.	TU Berlin	29	34
9.	Universität Stuttgart	27	39
10.	Universität Heidelberg	26	22
11.	Universität Bochum	26	15
12.	Universität Köln	25	8
13.	Universität München	23	20
14.	TU Braunschweig	19	21
15.	Universität Tübingen	19	19
16.	Universität des Saarlandes	19	9
17.	Universität Freiburg	16	12

* Anmeldungen am DDR-Patentamt

Tabelle 2

Der direkte Vergleich der Gesamtzahlen für einzelne Hochschulen wird insbesondere dadurch erschwert, dass jeweils unterschiedliche Fachbereiche involviert sind. Im Falle von Aachen dominieren der Maschinenbau und die übrigen Ingenieurwissenschaften bei den Patentanmel-

dungen, während bei der TU München die Chemie, aber auch die Biologie und die Agrarwissenschaften eine maßgebliche Rolle spielen (Tabelle 3). Im Falle der Universität Karlsruhe liegt die Elektrotechnik an erster Stelle, bei der Universitäten Göttingen oder Heidelberg die Chemie und Medizin (Becher et al. 1996: 200 f). Ranglisten von Hochschulen ohne eine weitere fachliche Differenzierung sind daher wenig aussagekräftig.

Patentanmeldungen	
RWTH Aachen	
Maschinenbau/Verfahrenstechnik	21
Übrige Ingenieurwissenschaften	12
Chemie	10
Bauingenieurwesen	6
Human/Veterinärmedizin	3
Biologie	3
Physik	3
Elektrotechnik	1
TU München	
Chemie	15
Maschinenbau/Verfahrenstechnik	11
Biologie	7
Agrar/Forst/Ernährungswissenschaften	6
Human/Veterinärmedizin	3
Bauingenieurwesen	3
Physik	2
Elektrotechnik	1

Tabelle 3

Bei der Beurteilung von Anmeldestatistiken ist neben der fachlichen Differenzierung in Rechnung zu stellen, dass bei einzelnen Hochschullehrern erhebliche Unterschiede in der Patentaktivität vorliegen. So melden 5 Prozent der patentaktiven Hochschullehrer 30 Prozent der Hochschulpatente an, wobei sie im Zeitraum 1990 bis 1997 jeweils mehr als 20 Anmeldungen tätigten (Abbildung 2; Krause/Schroll 1999: 16). Die Gruppe der weniger patentaktiven Hochschul-

lehrer mit ein bis zwei Anmeldungen umfasst dagegen 53 Prozent der Hochschullehrer und nur 18 Prozent der Anmeldungen. Das Patentgeschehen an Hochschulen wird somit maßgeblich von einer kleinen Gruppe sehr patentaktiver Hochschullehrer geprägt. Dieser Befund einer schiefen Verteilung (Lorenz-Kurve) ist allerdings nicht ungewöhnlich und in ähnlicher Weise auch bei Publikationen und deren Zitierung zu finden (Seglen 1992). Für die empirische Analyse von Hochschulanmeldungen bedeutet dieses Resultat, dass die Patentaktivitäten von Fachbereichen an Hochschulen maßgeblich durch den Ein- oder Austritt einzelner Hochschullehrer beeinflusst werden kann.

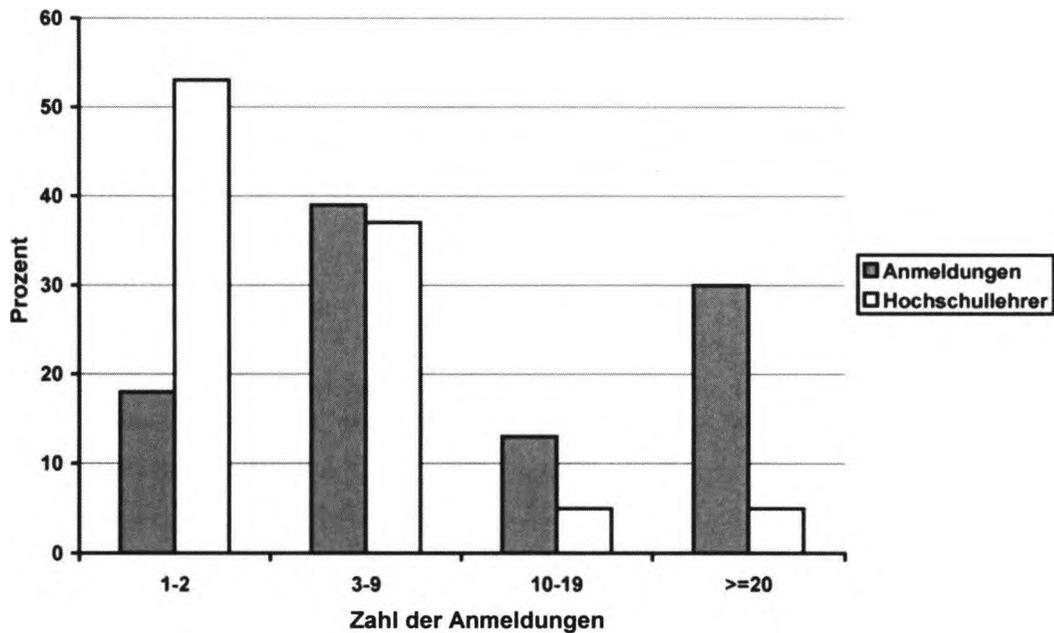


Abbildung 2

3. Schlussfolgerungen zur Aussagefähigkeit von Patentanalysen

Die Erhebung zu Patentanmeldungen aus deutschen Hochschulen zeigt, dass eine große Stichprobe vorliegt, die zu einer aggregierten Auswertung auf nationaler Ebene, aufgeschlüsselt nach Fachbereichen, geeignet ist. Die entsprechende Analyse führt insbesondere zu relevanten Aussagen zur Rolle von Hochschulen im deutschen Innovationssystem. Beim weiteren Herabbrechen auf die Ebene einzelner Fachbereiche treten jedoch Probleme der statistischen Aussagefähigkeit auf. Es besteht eine hohe Abhängigkeit von singulären Konstellationen wie der Patentierung von Hochschullehrern, dem Anfall patentrelevanter Ergebnisse oder der Existenz geeigneter Verwertungspartner. Das Ergebnis wird darüber hinaus maßgeblich vom „Patentklima“ an Hochschulen beeinflusst, bei dem vor allem Unterschiede zwischen allgemeinen gegenüber technischen Universitäten festzustellen sind. Entscheidend für die Anmeldung ist schließlich auch die Existenz oder das Fehlen einer Patent- oder Lizenzstelle an der jeweiligen Hochschule. Im Ergebnis bedeutet dies, dass die Stichproben auf der Ebene von Fachbereichen und

Instituten an Hochschulen klein und im Zeitverlauf instabil sind. Dabei ist zu bedenken, dass die aktuelle Zahl von 1500 Patentanmeldungen pro Jahr zwar erheblich ist, bei der Disaggregation auf die Vielzahl der deutschen Hochschulen aber zu sehr kleinen Teilmengen führt. Insofern empfiehlt es sich, sich bei der Analyse auf patentaktive Fachbereiche wie Maschinenbau, Chemie, Medizin oder Elektrotechnik zu konzentrieren und Stichproben für mehrjährige Beobachtungszeiträume zu bilden. Im übrigen ist zu bedenken, dass – zumindest in der gegenwärtigen Situation – die Patentierneigung eines Hochschullehrer einen erheblichen Einfluss hat. Von daher können hohe Patentzahlen an einem Fachbereich durchaus als Indikator für dessen Umsetzungs- und Praxisorientierung gewertet werden. Umgekehrt sind aber niedrige Patentzahlen kein zuverlässiger Indikator für fehlende Praxisorientierung, da die involvierten Hochschullehrer beispielsweise aus prinzipiellen Gründen die Privatisierung von Forschungsergebnissen ablehnen, die mit öffentlichen Mitteln erarbeitet worden sind.

4. Ausblick

Aufgrund der ständig steigenden Zahl von Patentanmeldungen gibt es gegenwärtig intensive politische Bemühungen, in allen Bundesländern Patent- und Lizenzstellen für Hochschulen einzurichten (FAZ vom 17.10.2000; BMBF 2000). Damit verbunden sind Planungen, das so genannte Hochschullehrerprivileg abzuschaffen oder zumindest in seinem Umfang erheblich zu reduzieren (Schmoch 2000a: 98 f). Sollte dies umgesetzt werden, müsste es an Hochschulen in Analogie zu privaten Unternehmen eine allgemeine Pflicht der Erfindungsmeldung geben, wonach die Hochschule über die Verwertung oder die Freigabe entscheiden kann. Damit dürften in wenigen Jahren bundesweite Statistiken zu Erfindungsmeldungen an Hochschulen vorliegen, die neben Hochschullehrern auch wissenschaftliche Mitarbeiter und anderes Hochschulpersonal umfassen, womit die Datengrundlage für statistische Analysen von Hochschulerfindungen wesentlich verbessert würde.

In der gegenwärtigen Situation ist eine Auswertung von Hochschulanmeldungen bei Beachtung der oben beschriebenen methodischen Grenzen zwar möglich, aber in der praktischen Umsetzung sehr aufwendig. Es gibt den eingangs beschriebenen Weg einer Identifikation relevanter Patentanmeldungen über den Professoren-Titel, wobei aber eine manuelle Nachbearbeitung für genauere Auswertungen erforderlich ist. Umgekehrt kann auf der Basis von Hochschul-Personenverzeichnissen in jedem Einzelfall geprüft werden, ob Patentanmeldungen vorliegen. Damit ist zwar auch das Personal ohne Professoren-Titel eingeschlossen; der Ansatz ist aber wesentlich aufwendiger.³

Vor diesem Hintergrund muss überlegt werden, ob es nicht einfachere Indikatoren für die Anwendungsorientierung der Hochschulforschung gibt. Ein möglicher Weg wäre hier die Auswertung von Fachpublikationen und deren Klassifizierung nach Anwendungs- oder Grundlagenorientierung. Der Vorteil solcher Analysen besteht darin, dass die verwertbaren Stichproben wesentlich größer als im Falle von Patentanmeldungen sind. So werden in der Datenbank Science Citation Index beispielsweise jährlich etwa 40.000 Fachpublikationen von

deutschen Hochschulen registriert. Die Publikationen lassen sich dabei über den disziplinären Kontext in techniknahe und grundlagenorientierte Publikationen einteilen, deren Relation durchaus zur Charakterisierung von Forschungseinrichtungen geeignet ist (Schmoch 2000b: 35). Außerdem ist es auch möglich, die Zeitschriften, in denen die Publikationen erscheinen, nach Anwendungsorientierung zu klassifizieren (Noyons et al. 1994). Schließlich könnten auch die Hochschullehrer nach der Anwendungsorientierung ihrer Forschung befragt werden (Schmoch 2000c: 104 ff; Rammer/Cernitzky 2000: 38 ff). Aufschlussreiche Hinweise zur Anwendungsorientierung von Instituten und Fachbereichen geben weiterhin Daten über die Finanzierungsstrukturen (Schmoch 2000c: 104 ff, Rammer/Carnitzky 2000: 38 ff). Zu unterscheiden ist hier insbesondere, in welchem Ausmaß die Finanzierung über Grundmittel, grundlagenorientierte Einrichtungen (z. B. Deutsche Forschungsgemeinschaft, Volkswagenstiftung), anwendungsorientierte Einrichtungen (z. B. Bundesministerium für Bildung und Forschung, Europäische Kommission) oder Unternehmen erfolgt. Methodisch problematisch ist hier bei den gegenwärtigen Strukturen, dass das Rechnungswesen an Hochschulen unterschiedlich organisiert ist und dass viele Hochschullehrer ein begrenztes Bewusstsein der für sie relevanten Kosten haben (Wittig et al. 2000; Abramson et al. 1997: 280 ff). Bei einer Befragung von Hochschullehrern zur Kostenstruktur ist nicht immer klar, inwieweit Gemeinkosten berücksichtigt werden.⁴

Insgesamt sind Patentanmeldungen als Indikator für die Anwendungsorientierung von Hochschulen grundsätzlich geeignet, wobei aber bei der Disaggregation auf der Ebene von Fachbereichen und einzelnen Einrichtungen enge methodische Grenzen gesetzt sind. Da die Erfassung von Patentanmeldungen an Hochschulen mit einem erheblichen Aufwand verbunden ist, empfiehlt es sich in der derzeitigen Situation, alternative Ansätze wie die spezifische Auswertung von Publikationsdaten oder die Befragung von Hochschullehrern zu verfolgen. Mit der Änderung der rechtlichen Rahmenbedingungen zum Patentwesen an Hochschulen könnten sich aber in wenigen Jahren Voraussetzungen ergeben, die eine statistische Analyse von Erfindungen aus Hochschulen erheblich vereinfachen.

Literatur

Abramson, H. N., Encarnação, J., Reid, P. P., Schmoch, U. (Hrsg.) *Technology Transfer Systems in the United States and Germany. Lessons and Perspectives*. National Academy Press. Washington, D. C. 1997a.

Bierhals, R., Schmoch, U. „Wissens- und Technologietransfer an Hochschulen“. Hrsg. Schmoch, U., Licht, G., Reinhard, M. *Wissens- und Technologietransfer in Deutschland*. Fraunhofer IRB Verlag. Stuttgart 2000. 74-89.

BMBF. *Pressemitteilung* "Änderung des Hochschullehrerprivilegs soll jetzt zügig angegangen werden", Berlin 1.11.2000.

³ Dieses Verfahren ist für die Fachbereiche Chemie und Maschinenbau zur Erstellung des CD-ROM-Studienführers 2000 eingesetzt worden.

⁴ Allerdings lassen sich durch eine geeignete Fragestellung vergleichbare Datengrundlagen erreichen.

Deutsches Patent- und Markenamt (DPMA) (Hrsg.). *Taschenbuch des gewerblichen Rechtsschutzes*. Karl Heymanns Verlag. Köln/München 2000

Felt, U., Nowotny, H., Taschwer, K. *Wissenschaftsforschung. Eine Einführung*. Campus. Frankfurt a.M./New York 1995.

Fuhrmann, F. U., Dzwonnek, D., Kleinbeck, U., Rieger, B. „Wie verteile ich meine Mittel? Die Suche nach quantitativen und qualitativen Kriterien“ Stifterverband für die Deutsche Wissenschaft (Hrsg.). *Reform Universitäten. Leistungsfähigkeit durch Eigenverantwortung*. Stifterverband. Essen 2000. 79-80.

Ganz-Brown, C. „Patent policies to fine tune commercialisation of government-sponsored university research“. *Science and Public Policy* Vol. 26, No. 6. 1999. 403-414.

Grupp, H., Schmoch, U. *Wissenschaftsbindung der Technik. Panorama der internationalen Entwicklung und sektorales Tableau für Deutschland*. Physica Verlag. Heidelberg 1992.

Krause, M., Schroll, M. *Deutsche Hochschullehrer zwischen Humboldtschem Ideal und Entrepreneurship. Patentierung und Verwertung von Forschungsergebnissen*. Institut für Angewandte Innovationsforschung. Bochum 1999.

Mazzoleni, R., Nelson, R.R. „The benefits and costs of strong patent protection: a contribution to the current debate“. *Research Polic*, Vol. 27. 1998. 273-284.

Merton, R.K. „Die normative Struktur der Wissenschaft“. Hrsg. Merton, R.K. *Entwicklung und Wandel von Forschungsinteressen. Aufsätze zur Wissenschaftssoziologie*. Suhrkamp Verlag. Frankfurt a. M. 1985. 86-99.

Noyons, E. C. M., Van Raan, A. F. J., Grupp, H., Schmoch, U. „Exploring the Science and Technology Interface: Inventor-Author Relations in Laser Medicine Research“. *Research Policy* Vol. 23 1994. 443-457.

Pavitt, K. “Do patents reflect the useful research output of universities?” *Research Evaluation* Vol. 7, No. 2. 1998. 105-111.

Rammer, C., Carnitzki, D. „Interaktion zwischen Wissenschaft und Wirtschaft – die Situation an öffentlichen Forschungseinrichtungen in Deutschland“. Hrsg. Schmoch, U., Licht, G., Reinhard, M. *Wissens- und Technologietransfer in Deutschland: Fraunhofer IRB Verlag*. Stuttgart 2000. 38-73.

Schmoch, U. *Wettbewerbsvorsprung durch Patentinformation. Handbuch für die Recherchepraxis*. Verlag TÜV Rheinland. Köln 1990.

Schmoch, U. „Indicators and the relations between science and technology“. *Scientometrics* Vol. 38, No. 1. 1997. 103-116.

Schmoch, U. „Rechtliche Situation von Patenten an Hochschulen“. Hrsg. Schmoch, U., Licht, G., Reinhard, M. *Wissens- und Technologietransfer in Deutschland.. Fraunhofer IRB Verlag*. Stuttgart 2000a. 96-103.

Schmoch, U. „Wissens- und Technologietransfer us öffentlichen Einrichtungen im Spiegel von Patent- und Publikationsstatistiken“. Hrsg. Schmoch, U., Licht, G., Reinhard, M. *Wissens- und Technologietransfer in Deutschland.*: Fraunhofer IRB Verlag. Stuttgart 2000b. 17-37.

Schmoch, U. „Ergebnis einer Umfrage bei technologienahen Forschungseinheiten an Universitäten“. Hrsg. Schmoch, U., Licht, G., Reinhard, M. *Wissens- und Technologietransfer in Deutschland.* Fraunhofer IRB Verlag. Stuttgart 2000c. 104-114.

Schmoch, U., Grupp, H., Mannsbart, W., Schwitalla, B. *Technikprognosen mit Patentindikatoren.* Verlag TÜV Rheinland. Köln 1988.

Seglen, P. O. „Towards a theory of citation: The skewed distribution of citedness“ Hrsg. Van Raan, A. F. J. et al. *Science and Technology in a Policy Context.* DSWO Press. Leiden 1992. 80-94.

Stankiewicz, R. *Academics and Entrepreneurs. Developing University-Industry Relations.* Frances Pinter Publishers. London 1986.

Stifterverband für die Deutsche Wissenschaft (Hrsg.) *Reform Universitäten. Leistungsfähigkeit durch Eigenverantwortung.* Stifterverband. Essen 2000.

Wittig, C., Fischer, A., Ebers, N., Wechselbaumer, J. „Was kosten wir, was leisten wir? Argumente für ein neues Rechnungswesen“ Hrsg. Stifterverband für die Deutsche Wissenschaft *Reform Universitäten. Leistungsfähigkeit durch Eigenverantwortung.* Essen 2000: Stifterverband. 81-82.

Tabellen und Abbildungen

Abbildung 1: Patentanmeldungen aus deutschen Hochschulen am Deutschen Patent- und Markenamt (DPMA)

Quelle: PATDPA, eigene Erhebungen.

Abbildung 2: Patentanmeldungen aus deutschen Hochschulen nach Anteil der Anmeldungen und der Hochschullehrer nach Größenklassen, 1990-97 (Stichprobe: 743 Anmeldungen)

Quelle: Krause/Schroll (1999: 16)

Tabelle 1: Fachliche Zugehörigkeit von Hochschul-Professoren mit Patentanmeldungen im Jahr 1993

Quelle: PATDPA, eigene Erhebungen

Tabelle 2: Hochschulen mit hohem Patentaufkommen in den Jahren 1985 und 1993

Quelle: Becher et al. (1996: 39)

Tabelle 3: Patentanmeldungen ausgewählte Hochschulen im Jahr 1993 nach Fachbereichen der Erfindungsherkunft

Quelle: Becher et al. (1996: 200)

ISA: Informationssystem Studienwahl & Arbeitsmarkt

Michael Weegen

In der gegenwärtigen bildungspolitischen Diskussion wird dem Studienwahlverhalten viel zu wenig Beachtung geschenkt. Analysen des Wahlverhaltens von Schülern bzw. Studienberechtigten zeigen für die unterschiedlichsten Studiengänge ein nahezu konstantes Verhaltensmuster auf: Die individuellen Entscheidungen verlaufen, sofern sie mit Blick auf den Arbeitsmarkt erfolgen, prozyklisch. Die Studienberechtigten orientieren sich dabei – in Ermangelung geeigneter Informationen - in der Regel an den ihnen zugänglichen aktuellen Arbeitsmarktdaten, ohne deren erwartbare Entwicklung bis zu dem Zeitpunkt, an dem sie in das Beschäftigungssystem eintreten werden, zu antizipieren. Dieses prozyklische Studienwahlverhalten wurde in den vergangenen Jahren regelmäßig beobachtet und beklagt, so z. B. mit Blick auf die Entwicklung des Teilarbeitsmarktes Schule. Aktuell lässt dieses Verhaltensmuster sich sehr deutlich im Bereich der ingenieur- und naturwissenschaftlichen Fachrichtungen beobachten.

Für die Studierenden selbst - aber auch für die Planungen in den Hochschulen und für die Wissenschaftsverwaltungen - stellt dieses prozyklische Studienverhalten ein erhebliches Problem dar, da es für Individuen das Arbeitslosigkeitsrisiko erhöht, da es für die Gesellschaft insgesamt ohnedies bestehende Ungleichgewichte auf dem Arbeitsmarkt verstärkt und da es in den Hochschulen zu Fehlsteuerungen beim Ressourceneinsatz beitragen kann. Und letztendlich hat es dazu beigetragen, dass sich die Erfolgsquote insbesondere an den Universitäten unaufhaltsam vermindert hat. Rechnet man das Aufkommen der Studiengangwechsler mit, kann man derzeit für die Universitäten insgesamt davon ausgehen – bei sehr starken fachspezifischen Unterschieden - dass jeder zweite Anfänger an Universitäten sein ursprüngliches Ziel (Abschluss laut Erstimmatrikulation) nicht erreicht, wobei arbeitsmarktbezogene Studienabbrüche sicherlich eine erhebliche Bedeutung erlangt haben (Hörner 1998: 71). Von 100 Studierenden, die sich in ein erstes Fachsemester an einer Universität einschreiben, erreichen etwa 60 in den Lehramtsstudiengängen, ebenfalls 60 bei den Maschinenbauern und Juristen, 50 bei den Elektrotechnikern und Physikern, 40 bei den Wirtschaftswissenschaftlern und Informatikern, 30 bei den Mathematikern, 20 bei den Germanisten und 10 bei den Philosophen ihren entsprechenden Abschluß.

Freilich ist es unstrittig, dass die Gruppe derer, die die Universitäten ohne Abschluss verlassen, hinsichtlich ihrer Motive sehr heterogen ist (Klemm /Weegen 2000).

Derjenige, der heute eine Studienwahlentscheidung trifft, will in der Regel Informationen über die Situation im Hochschulbereich und zu zukünftigen Berufsaussichten mit einbeziehen, wiewohl sicherlich mit unterschiedlichem Gewicht. Wie bereits schon erwähnt, erfolgt der Blick auf die gegenwärtige Situation des oder der relevanten Teilarbeitsmarktes/-märkte mit dem Wunsch, gute Aussichten auf einen möglichst sicheren Arbeitsplatz zu haben (AG Hochschulforschung 1998: 30). Weitergehende Informationen gehören in den Bereich der Spezifika, sofern sich das Fach und der Teilarbeitsmarkt dann überhaupt noch genügend scharf abgrenzen lassen. Das Interesse bezogen auf die Fächer und den Hochschulbereich insgesamt richtet sich primär auf die Studienbedingungen und weniger auf die Nachfrageentwicklung,

zudem hier abrufbare Informationen fehlen. Veröffentlichungen wie der von Stern und CHE gemeinsam herausgegebene Studienführer kommen dem Informationsbedürfnis nach Studienbedingungen mittlerweile recht professionell nach.

Im Hinblick auf naturwissenschaftlich – technische Fächer ist es aber beispielsweise nahezu unmöglich geworden, deren Absolventen bestimmten Teilarbeitsmärkten - geschweige denn Berufsfeldern analog eines Verteilungsschlüssels - zuzuordnen. Diese Inkompatibilität ist vor dem Hintergrund eines anhaltenden Strukturwandels des Arbeitsmarktes zu verstehen, der sich seit Mitte der siebziger Jahre mit wachsender Rasanzen vollzieht. Deshalb ist es nicht verwunderlich, dass es der Arbeitsmarkt- und Berufsforschung schon seit Jahren nicht mehr gelingt, geschlossene Berufsklassifikationen zu erstellen, die den tatsächlichen akademischen Tätigkeiten, Funktionen und Berufsfeldern gerecht werden (Minks 1999). Als Konsequenz daraus hat man mit Blick auf bestimmte Fächer, Fächergruppen und Abschlussprofile davon Abstand genommen, direkte fach- und berufsbezogene Vorausschätzungen zu erstellen (ebenda). Die klassische Bedarfsprognose ist für viele Fächer passé. Den Blick aber ausschließlich nur auf die Situation der großen Sektoren (primärer, sekundärer und tertiärer Sektor, BLK 1995) zu richten, kann nicht der Weisheit letzter Schluss sein.

Denn an aussagekräftigen Parametern zur zukünftigen Einschätzung des Arbeitsmarktes nach Fächern und Fachgebieten mangelt es eigentlich nicht.

Die Aufgabe und Zielsetzung von ISA besteht nun darin, die Informationen und Daten zu einzelnen Studienbereichen und Teilarbeitsmärkten - soweit dies strukturell vertretbar ist - aufeinander abzustimmen und anschließend übersichtlich und adressatengerecht im Internet zu präsentieren, um ein hohes Maß an Orientierungsmöglichkeiten zu geben. Um die Aussagen zu einzelnen Studienbereichen fortschreibungsfähig zu machen und auf eine seriöse datengestützte Grundlage zu stellen, sind deshalb Grunddaten aus drei **statistischen Welten** zusammengetragen und anschließend nach einer vertretbaren Systematik aufeinander abgestimmt worden:

- Hochschul- bzw. Studierendenstatistik - mit den Daten zu Studienanfängern, Studierenden, Absolventen, Studiendauer, Alter, Studienerfolg (Quelle: Statistisches Bundesamt).
- Arbeitsmarktstatistik bzw. Erwerbstätigenstatistik - mit Angaben zur Erwerbstätigkeit von Akademikern einschl. ihrer Altersstruktur (Quelle: Institut für Arbeitsmarkt- und Berufsforschung bzw. Statistisches Bundesamt - Mikrozensusdaten).
- Arbeitslosenstatistik - mit Angaben nach Qualifikationen, Abschlüssen und Alter der beim Arbeitsamt gemeldeten Personen (Quelle: Bundesanstalt für Arbeit).

Die Auswertung hat sich auf die 26 am meisten nachgefragten Studienbereiche erstreckt. Zusätzlich mit Blick auf die Teilarbeitsmärkte sind die Bereiche Schule/Lehrerbildung und Informations- und Kommunikationstechnologie/ Multimedia dargestellt worden.

Die Abstimmung der Datensätze zu Teilarbeitsmärkten und Studienbereichen kann aber - und das ist bisher ein zentrales Problem der Hochschul- und Arbeitsmarktstatistik gewesen - nicht automatisch erfolgen. Zum einen sind in den Sektoren Hochschulstatistik, Mikrozensusdaten/ Erwerbstätigenentwicklung und Arbeitslosenstatistik unterschiedliche Systematiken und

Kennziffern vorhanden. Zum anderen sind die Konturen der in Frage kommenden Teilarbeitsmärkte in bezug auf die gewählten Fächergruppen höchst unterschiedlich.

Die Entwicklung von Studienanfängern, Absolventen, Erwerbstätigenzahlen und Angaben über die Altersstruktur von Beschäftigten können aber bei Wahl der richtigen Tiefenschärfe - in Bezug auf die zusammengestellten Daten - perspektivisch Aussagen über die zukünftige Situation von denjenigen zulassen, die heute eine Studienwahlentscheidung treffen.

Mit Blick auf die spätere Arbeitsmarktsituation können diese Daten sicherlich nicht hellseherisch wirken. Treffsichere Prognosen oder selbst nur verbindliche Empfehlungen sind aufgrund der Dynamik des Arbeitsmarktes nahezu unmöglich geworden. Diese Unsicherheit gilt immer stärker auch für die Beschäftigtenentwicklung und die Arbeitsmarktchancen im öffentlichen Sektor. Von daher verbietet sich eine Perspektive die beispielsweise ausschließlich rein auf eine statische Ersatzbedarfsberechnung des Arbeitskräfteangebotes verengt wird.

So kann eine kritische Abwägung der Informationen durchaus nahelegen, antizyklisch zu studieren – also gegen den gegenwärtigen Trend eine Studienwahlentscheidung zu treffen. Oder aber sie lässt bei ganz bestimmten Studienbereichen die Schlussfolgerung zu - wenn sozusagen alle Parameter in eine negative Richtung weisen, sich eine mögliche Entscheidung in dem ausgesuchten Bereich bei Anbetracht der eindeutigen Perspektiven gut zu überlegen. Insgesamt lassen sich theseartig folgende Zielsetzungen formulieren:

- Informationen über Projektionen und Rahmendaten zu Teilarbeitsmärkten mit der Zielsetzung einer antizyklischen Betrachtung - zwecks Abwägung der eigenen Studienwahlentscheidung;
- Informationen über Studienerfolg, -fachwechsel, -abbruch und weitere Bedingungen;
- Verweise auf Informationen über Studienangebote und ggf. -bedingungen;
- Bündelung und Vernetzung von vorhandenen Informationen;
- Beitrag zum schonenden Umgang finanzieller Ressourcen von Studierenden, Hochschulen und Staat.

Wie stellt sich die Vernetzung von Informationen im einzelnen dar?

Am Beispiel des Studienbereichs Architektur soll eine Darstellung konkret vorgestellt werden (unter www.uni-essen.de/isa sind für die am meisten nachgefragten Studienbereiche aktuelle Informationen abrufbar).

Architektur

Hochschule / Arbeitsmarkt

Das Studium der Architektur kann sowohl an Universitäten / Technischen Hochschulen als auch an Fachhochschulen aufgenommen werden. Die vier verschiedenen Aufgabenfelder dieses Studienbereichs sind

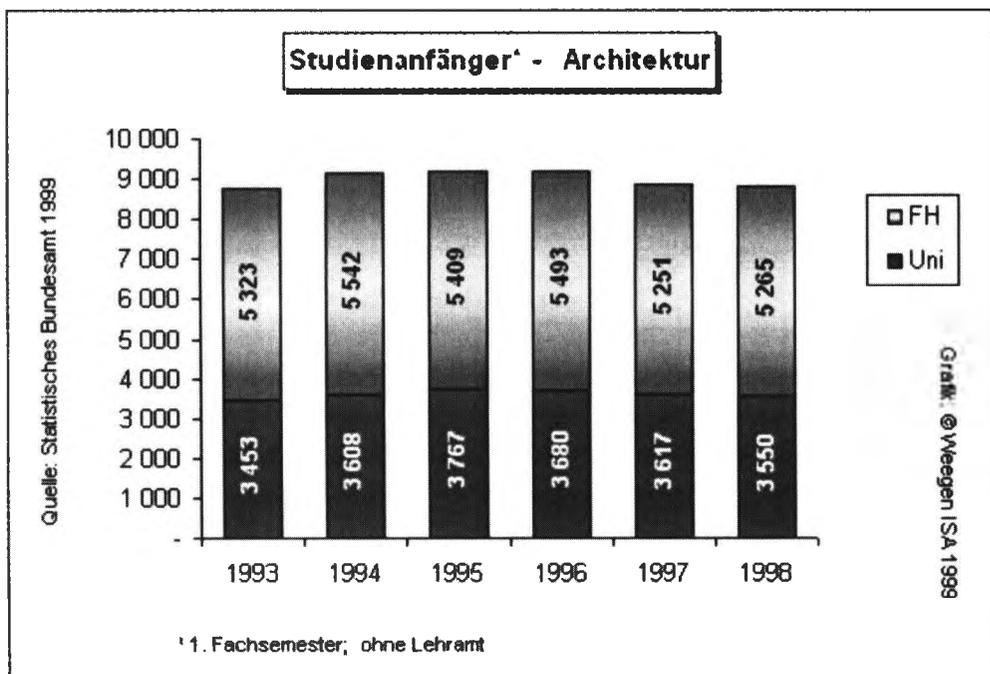
- Architektur (Hochbau)
- Innenarchitektur (Innenräume)
- Landschaftsarchitektur (Garten- und Landschaftsplanung)
- Stadtplanung (Stadt-, Regional- und Raumplanung)

Als wesentliche Grundlagenfächer des Studienbereichs Architektur lassen sich nennen: Kunstgeschichte, Baugeschichte, künstlerisches und räumliches Gestalten, Baustoffkunde, Städtebau und Vermessungskunde. Dieses Angebotsspektrum gilt für Universitäten und Fachhochschulen gleichermaßen, allerdings mit unterschiedlichen Gewichtungen. Im Hauptstudium legen sich die Studierenden mit einer Schwerpunktsetzung fest, die mit Blick auf die spätere Berufstätigkeit ausgerichtet ist und je nach Aufgabenfeld aus den Fachgebieten Städtebau, Denkmalpflege, Gebäudeentwurf, Bauphysik oder Baukonstruktion gewählt wird. Wer sich einen inhaltlichen Einblick über das Architekturstudium an einer Fachhochschule machen will, sei auf das Angebot der Fachhochschule Darmstadt verwiesen. Von den Universitäten bzw. Technischen Hochschulen sei hier das virtuelle Angebot der RWTH Aachen und der TU Berlin hervorgehoben. Mit Blick auf die Studienmöglichkeiten und Studiengänge empfiehlt sich der Studien- und Berufswahlführer, der gemeinsam von der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung und der Bundesanstalt für Arbeit herausgegeben wird. Einen umfassenden Überblick über die Aufgabenstruktur des Architekten und auch studienrelevanter Fragen bieten die Angebote der nordrhein-westfälischen Architektenkammer und des Hauptverbandes der deutschen Bauindustrie. Einen systematischen Überblick des Faches Architektur erhält man auf dem Dinoserver und bei ArchiNet der FH Bochum.

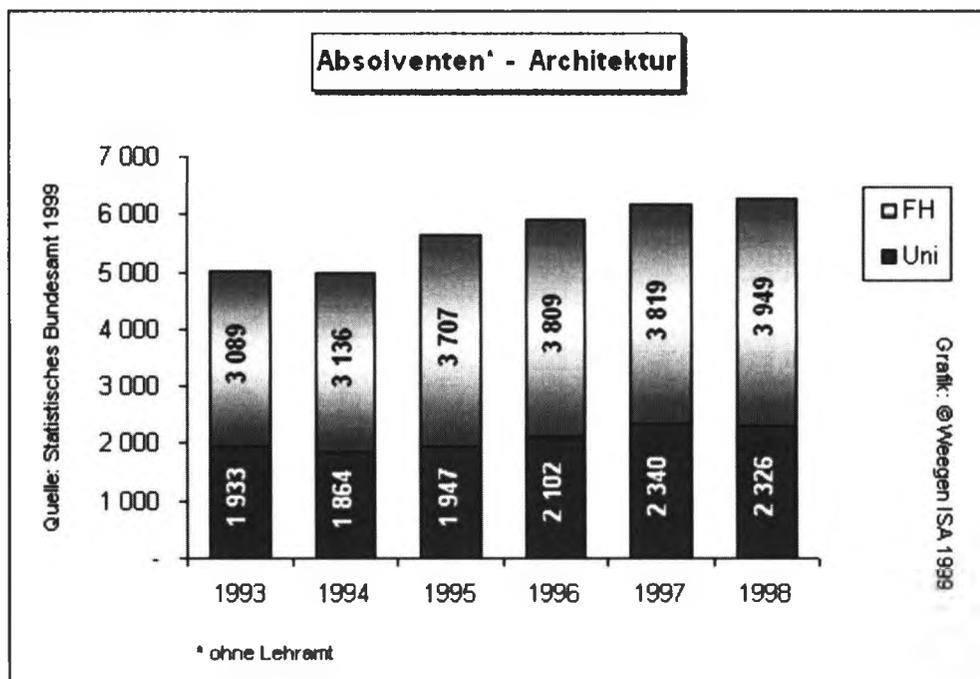
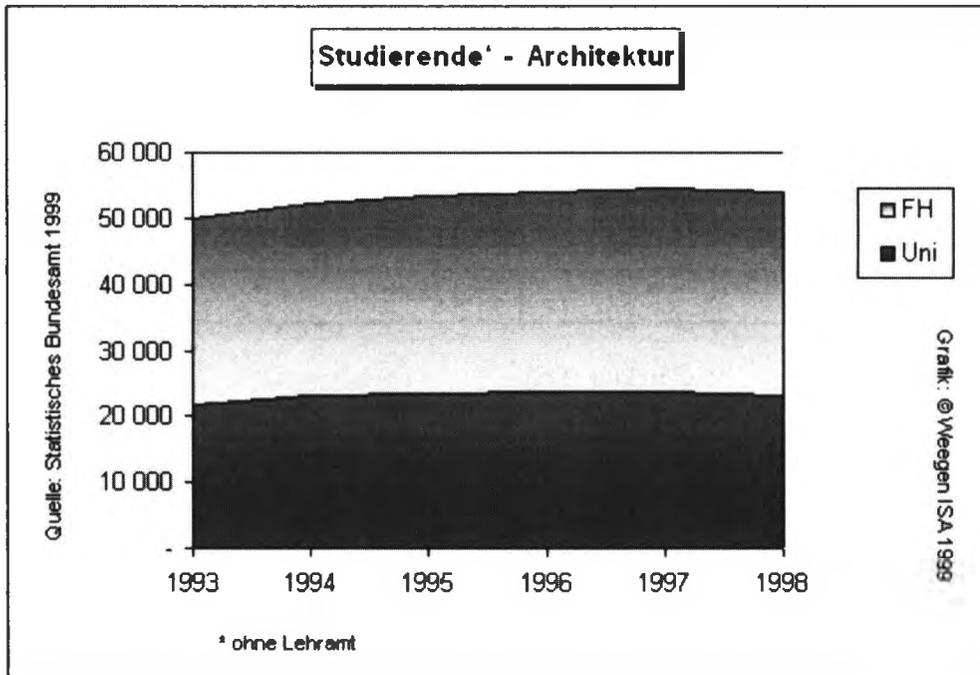
Sowohl die Studierenden- als auch die Studienanfängerzahlen in den Architekturstudiengängen an Universitäten und an Fachhochschulen sind seit 1993 auf sehr stabilem Niveau geblieben (Ungefähr 10 % der Studierenden haben davon das Fach Innenarchitektur belegt). In den alten Ländern besteht dieses Niveau seit Anfang der achtziger Jahre und ist auf die Zulassungsbeschränkung der Studiengänge zurückzuführen. Im WS 1999/2000 lag der NC-Notendurchschnitt im Rahmen des ZVS-Verfahrens je nach Bundesland zwischen 1,9 (Thüringen) und 2,6 (Schleswig - Holstein). Mehr über den aktuellen Bewerberstand und das Studienplatzangebot erfährt man auf dem Server der Zentralstelle für die Vergabe von Studienplätzen (ZVS).

Studiendauer und Alter - Architektur (Uni)	
Regelstudienzeit:	9 - 10 Semester (Uni)
Tatsächliche Studiendauer*:	13,4 Fachsemester (Uni)
Alter bei Abschluss*:	29,4 Jahre
Studiendauer und Alter - Architektur (FH)	
Regelstudienzeit:	9 Semester (FH)
Tatsächliche Studiendauer*:	11,1 Fachsemester (FH)
Alter bei Abschluss*:	29,0 Jahre

*) Von deutschen Erstabsolventen mit bestandener Prüfung erhobene Daten.
Quelle: Statistisches Bundesamt, Fachserie 11, Reihe 4.2 Prüfungen an Hochschulen 1997, Wiesbaden 1998)



Aufgrund der langjährigen Zulassungsbeschränkung sind die Absolventenzahlen relativ stabil geblieben. Hier ist der Anstieg seit 1996 ist auf die Einbeziehung der Studienplätze in den neuen Bundesländern zurückzuführen. Seit 1991 hat die Studienanfängerzahl bedingt durch die Wiedervereinigung um ca. 20% zugenommen.

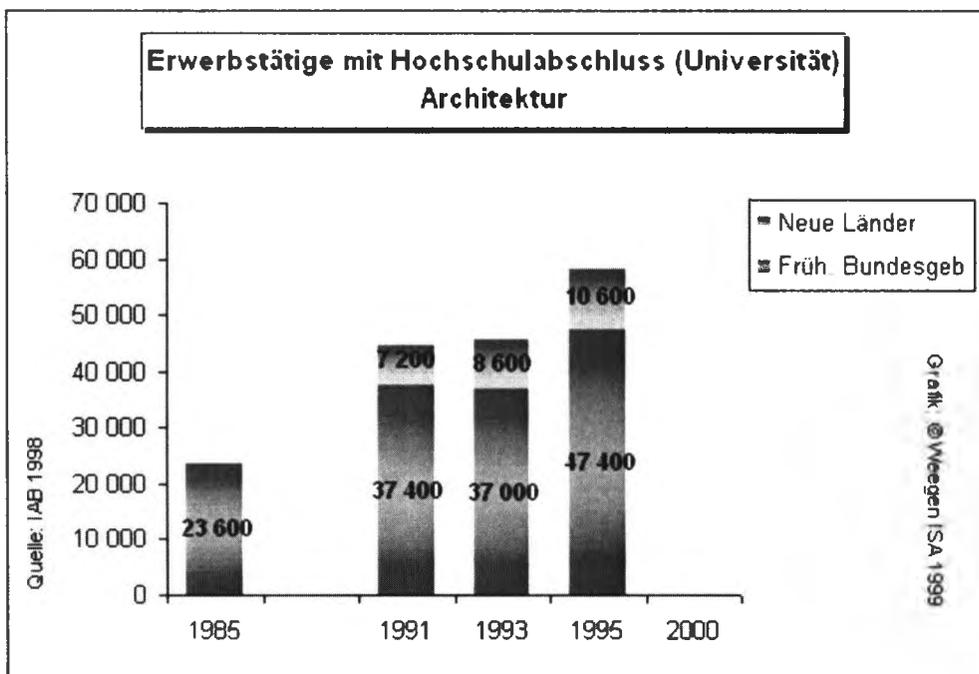


Dieser Anstieg spiegelt sich bei den Absolventenzahlen seit 1996 – proportional - wider. Setzt man das Niveau der Studienanfängerzahlen vom Jahr 1993 mit den entsprechenden Absolventenzahlen von 1997 bzw. 1998 in Beziehung, so lässt sich tendenziell die Aussage ableiten, dass bundesweit das Erfolgsniveau bei den Architekturstudiengängen an Fachhochschulen bei ca. 70% gelegen hat und damit vergleichsweise hoch ist. Nahezu gleich hoch ist das Erfolgsniveau aber auch an den Universitäten bzw. den Technischen Hochschulen gewesen. Bei Fortschreibung dieses hohen Niveaus ist in den nächsten fünf bis sechs Jahren jährlich mit ca. 6.000 bis 6.500 Absolventen zu rechnen, die einen Diplomabschluss in Architektur erworben haben werden, ca. 4.000 davon mit Fachhochschulabschluss.

Architektur

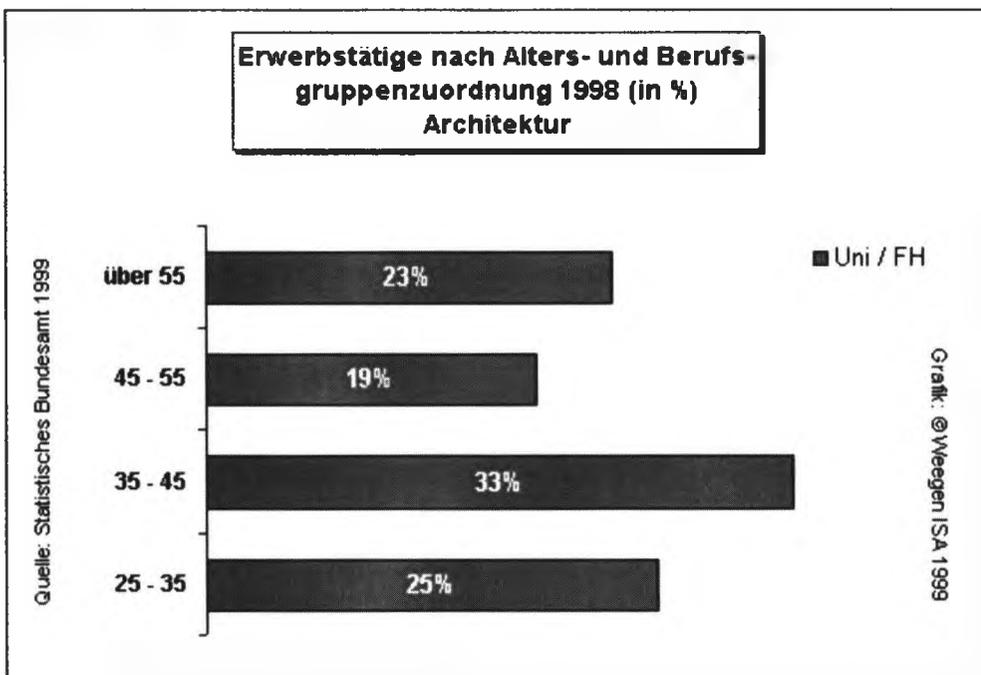
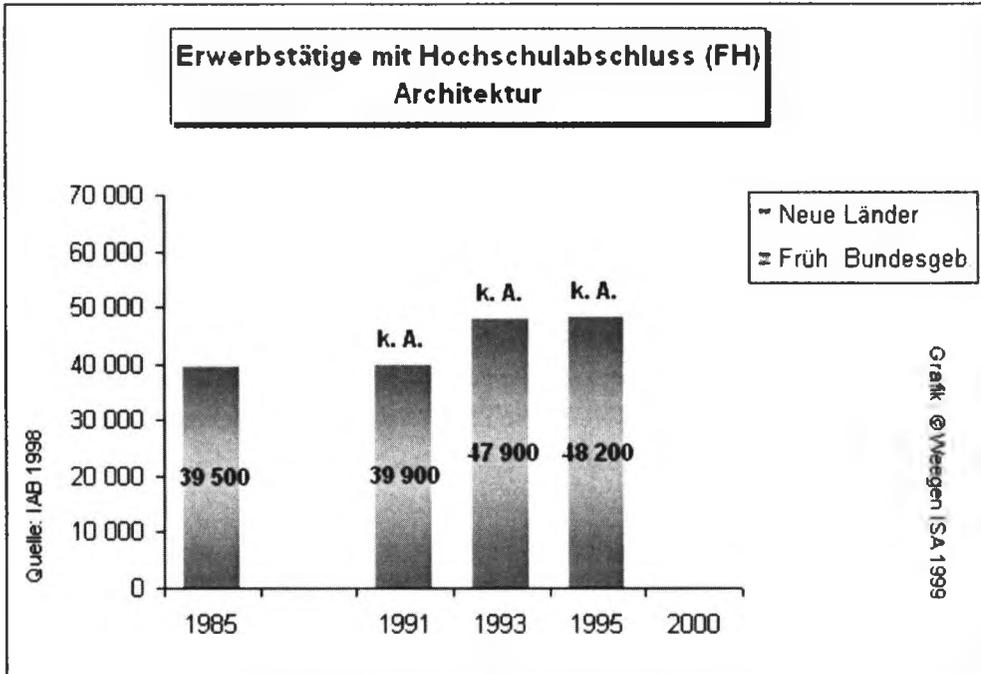
Hochschule / Arbeitsmarkt

Die Zahl der Erwerbstätigen mit einem universitären Diplomabschluss für Architektur hat sich laut IAB von 23.600 (1985) auf 47.400 (1995) in den alten Ländern kontinuierlich erhöht. Bei den Architekten mit Fachhochschulabschluss ist der Anstieg allerdings wesentlich flacher verlaufen: Hier nahm in den alten Ländern die Erwerbstätigenzahl von 39.500 auf 48.200 zu. Für die neuen Länder zeigt sich ausgehend vom Jahr 1991 bei den erwerbstätigen Architekten mit einem universitären Abschluss ein Zuwachs von 7.200 (1991) auf 10.600 (1995). Für den FH - Bereich liegen noch keine Angaben vor. Insgesamt gab es laut Angaben des Statistischen Bundesamtes im Jahr 1998 ca. 118.000 Architekten (einschl. vergleichbarer Abschlüsse) in Deutschland. Die IAB-Auswertung bezieht aus systematischen Gründen die erwerbstätigen Innenarchitekten nicht mit ein.



Die Teilzeitquote der erwerbstätigen Architekten mit Universitätsabschluss hat laut IAB in den alten Ländern 1995 bei 14% und bei den FH-Absolventen bei 11% gelegen. Laut Abschlussbezogener Auswertung waren 1995 ca. 26% aller Architekten älter als 50 Jahre.

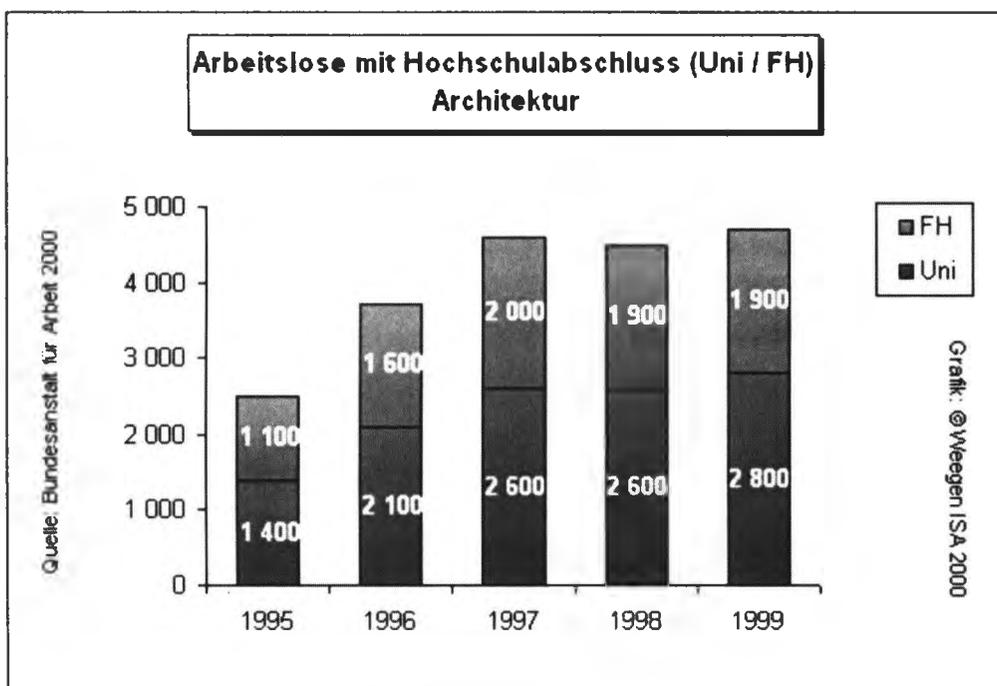
Die Altersstruktur weist berufsbezogen gemäß Statistischem Bundesamt bei den erwerbstätigen Architekten, die über 55 Jahre alt sind für 1998 ein Anteil von 23 % aus (Universitäts- und Fachhochschulabschluss). Bei einem Austrittsalter von durchschnittlich 65 Jahren hätte dies zur Folge, dass pro Jahr durchschnittlich ca. 2000 bis 2.500 Architekten ausscheiden würden. Somit wird in diesem Zeitraum die Zahl der Absolventen aus den Hochschulen (Uni und FH) mehr als doppelt so groß sein wie die Zahl der altersbedingten Austritte aus dem Erwerbsleben. Bei einer Verminderung des durchschnittlichen Austrittsalters auf 63 Jahre würde sich die Zahl der Austritte auf 3.500 pro Jahr erhöhen.



Aufgrund der derzeitigen Entwicklung der Baubranche bzw. der Einschätzung des Hauptverbandes der deutschen Baubranche und führender Wirtschaftsinstitute (DIW, RWI) ist in den nächsten Jahren nicht von weiteren gravierenden Verschlechterungen der Baubranche auszugehen. Eine Wachstumsannahme muss aber als optimistische Variante angesehen werden. Ein nennenswerter Mehrbedarf an einzustellenden Architekten ist unwahrscheinlich. Die Marktmöglichkeiten für freischaffende Architekten scheinen bereits gegenwärtig ihre Grenzen zu erreichen. Aus diesen Gründen fällt auch die Einschätzung der nordrhein-westfälischen Architektenkammer zu den künftigen Berufsaussichten sehr vorsichtig aus und weist dabei vorausschauend deutlich auf neue Rahmenbedingungen hin:

"Die berufliche Situation der Architekten ist einem starken Wandel unterworfen. Dies ergibt sich besonders aus der Entwicklung des gemeinsamen Marktes in der Europäischen Union. Traditionell bestehen in den europäischen Ländern sehr unterschiedliche Rahmenbedingungen für die Berufsgruppen. Dies gilt auch für die regionalen Märkte der Bauwirtschaft, so dass sich für die kommenden Jahren starke strukturelle Veränderungen ergeben werden. Im Vergleich der europäischen Länder unterscheidet sich die Zahl der Architekten in Bezug auf die Bevölkerung deutlich. Mit 1.000 Architekten pro 1 Mio. Einwohner hat Deutschland eine außergewöhnlich hohe Architektendichte gegenüber anderen Staaten, z. B. den Niederlanden mit nur 130 Architekten pro 1 Mio. Einwohner." (vgl. nordrhein-westfälische Architektenkammer 1999 <http://www.aknw.de/service/ak35.htm> vom 30.07.1999).

Derjenige, der heute vor der Entscheidung seiner Studienwahl steht, sollte sich über diesen bevorstehenden Wandel und die sich ändernden Rahmenbedingungen im klaren sein.



Die Zahl der arbeitslosen Architekten und Architektinnen hat sich seit 1995 fast verdoppelt. Angesichts der ungewissen Perspektiven am Bau waren zudem viele neue Positionen nur befristete Beschäftigungsverhältnisse oder basierten auf freier Mitarbeit. Laut Arbeitsmarktbeobachtung der ZAV konnten frisch diplomierte Architekten kaum auf andere Weise ihre ersten beruflichen Erfahrungen machen. Die aufkommende Anspannung auf dem Arbeitsmarkt hat sogar dazu geführt, dass erfahrene Architekten mitunter erhebliche Gehaltsabschläge hinnehmen mussten. Die wenigen Angebote mit Festanstellungen sind in der Regel mit hohen Anforderungen verbunden gewesen. Gefragt waren kreative Jungarchitekten unter 35 Jahre mit Erfahrungen bei der Ausschreibung, der Vergabe und Abrechnung von Bauprojekten. CAD-Kenntnisse werden vorausgesetzt (ZAV 1999). Eine kurze Beschreibung traditioneller und neuer Tätigkeitsfelder findet sich in der Broschüre der ZAV, Arbeitsmarktinformationen – Architektinnen und Architekten, Frankfurt 1999.

Literaturangaben

Arbeitsgruppe Hochschulforschung: Studiensituation und Studierende in den 90er Jahren, Datenalmanach 1993 – 1998, Konstanz 1998.

Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie: Grund- und Strukturdaten (fortlaufend), Bonn 2000.

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK): Beschäftigungsperspektiven der Absolventen des Bildungswesens. Materialien zur Bildungsplanung und zur Forschungsförderung, Heft 45, Bonn 1995.

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK) gemeinsam mit der Bundesanstalt für Arbeit: Studien- und Berufswahl 2000/01, Nürnberg 2000.

CHE (Centrum für Hochschulentwicklung): Test-Spezial: Studienführer Chemie & Wirtschaftswissenschaften, Mai 1998, erhältlich bei der Stiftung Warentest und Stern, Start und CHE fortlaufend Studienführer für Jura, Informatik, Mathematik, Physik, Wirtschaft, Chemie 1999 und Architektur, Bauingenieurwesen, Maschinenbau und Elektrotechnik 2000.

Hörner, W.: Studienerfolg, Studienabbruch und Studienkosten im internationalen Vergleich. In: Internationale Hochschulstatistik, Bd. 12 der Schriftenreihe Bundesstatistik, Wiesbaden 1998.

Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit: Materialien aus der Arbeitsmarkt- und Berufsforschung. Sonderserie Akademiker/innen – Studium und Arbeitsmarkt Nürnberg 1998.

Klemm, K./ Weegen, M.: Wie gewonnen so zerronnen. In: Rolff, H. G. u.a. (Hrsg.): Jahrbuch der Schulentwicklung – Band 11, Wiesbaden 2000.

Minks K.H.: Die meisten Hochschulen bilden falsch aus. In: Handelsblatt, 15./16. 01.1999.

Autoren

Dipl.-Wi.-Ing. Christian R. Bayer
Universität Karlsruhe
Institut für Wirtschaftstheorie & OR
76128 Karlsruhe
bayer@wior.uni-karlsruhe.de

Professor Dr. Hans-Dieter Daniel
Universität Zürich
Evaluationsstelle
Künstlergasse 15
CH-8001 Zürich

Jürgen Güdler
Geschäftsstelle der DFG
Kennedyallee 40
53175 Bonn
Juergen.Guedler@dfg.de

Dr. Stefan Hornbostel
Institut für Soziologie
Friedrich-Schiller-Universität Jena
Otto-Schott-Str. 41
07740 Jena
hornbostel@soziologie.uni-jena.de

Dr. habil. Uwe Jensen
Institut für Statistik und Ökonometrie
Universität Kiel
Olshausenstr. 40
24118 Kiel
jensen@stat-econ.uni-kiel.de

Professor Dr. Helmut Kromrey
Freie Universität Berlin, Institut für Soziologie
Babelsberger Str. 14-16
10715 Berlin
kromrey@zedat.fu-berlin.de

Dr. Ulrich Schmoch
Fraunhofer-Institut für Systemtechnik und
Innovationsforschung (ISI)
Breslauer Straße 48
D-76139 Karlsruhe
us@isi.fhg.de

Dr. Michael Weegen
Informationssystem Studienwahl und Arbeitsmarkt
(ISA)
Fachbereich 2
Universität GH Essen
Universitätsstr. 2
45117 Essen
michael.weegen@uni-essen.de

Das CHE Hochschulranking

Untersuchungen

1998

Uni: BWL, Wirtschaftswissenschaft/ Ökonomie, VWL, Wirtschaftspädagogik, Chemie, Biochemie, Lebensmittelchemie

FH: Betriebswirtschaft, VWL

1999

Uni: Informatik, Jura, Mathematik, Physik, Wirtschaftsinformatik

FH: Informatik, Wirtschaftsinformatik

2000

Uni: Maschinenbau / Verfahrenstechnik, Architektur, Bauingenieurwesen, Elektrotechnik, Wirtschaftsingenieurwesen

FH: Maschinenbau / Verfahrenstechnik, Architektur, Bauingenieurwesen, Elektrotechnik, Physikalische Technik, Vermessungswesen, Wirtschaftsingenieurwesen

2001

Uni: Anglistik/ Amerikanistik, Erziehungswissenschaften, Germanistik, Geschichte, Psychologie

2002

Uni: Jura, Politikwissenschaft, Sozialwissenschaften, BWL, Ökonomie, VWL, Wirtschaftsinformatik, Wirtschaftsingenieurwesen, Wirtschaftspädagogik

FH: Wirtschaftsrecht, Sozialwesen, Betriebswirtschaft, Wirtschaftsinformatik, Wirtschaftsingenieurwesen

Veröffentlichungen

test spezial Studienführer, Stiftung Warentest / CHE, 1998

start, Der Studienführer von stern, start und CHE, 1999

start, Der Studienführer von stern, start und CHE, 2000

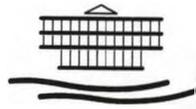
Der Studienführer 2001, stern spezial Campus & Karriere, 1/2001

Internet

www.dashochschulranking.de



Centrum für
Hochschulentwicklung



Verlag Bertelsmann Stiftung

Hochschulranking - Aussagefähigkeit, Methoden, Probleme

Seit Ende der achtziger Jahre die ersten Hochschulrankings in Deutschland erschienen, sind sie umstritten. Methodisch haben sie sich jedoch erheblich weiterentwickelt. Befragungsergebnisse von Studierenden wie Professoren fließen heute ebenso ein wie Fakten, basierend auf statistischen Angaben, Patentanalysen und bibliometrischen Messungen. Doch wie ist es bestellt um ihre Aussagekraft? Ist ihre Wirkung bei der Studienplatzwahl nachweisbar? Welche methodischen Fallstricke birgt der Vergleich hochschulischer Leistungen in Forschung und Lehre? Von der Analyse von Patentaufkommen und Drittmitteln zur **Beurteilung von Forschungsleistungen** über statistisch-methodische Neuerungen, Evaluation mittels Befragungen bis zur Information über den Arbeitsmarkt: Die in diesem Tagungsband zusammengefassten Beiträge spiegeln die breite Palette methodischer Herausforderungen bei der Erstellung fundierter Hochschulrankings. Beispielhaft steht in den meisten Beiträgen das umfassendste deutsche Ranking, das jährlich aktualisiert vom CHE (in Zusammenarbeit mit dem stern) veröffentlicht wird (www.dashochschulranking.de).

ISBN: 3-89206-575-5