

# Bessere Didaktik trotz schlechterer Rahmenbedingungen – Ergebnisse zweier Befragungen zur Situation von Lehre und Studium an der TU Dresden

*René Krempkow<sup>1</sup>*

*Institut für Soziologie der TU Dresden*

## 1. Einleitung

Viele Studiengänge an Hochschulen sind in den letzten Jahren in der Situation, bei knapper werdenden öffentlichen Mitteln und steigenden Studentenzahlen die Qualität ihrer Ausbildung sichern zu müssen. Hierbei erweist sich die vom Gesetzgeber vordefinierte Fokussierung von qualitätssichernden Maßnahmen auf die Qualität von Lehrveranstaltungen möglicherweise als problematisch, da die Rahmenbedingungen der Lehre bzw. die Studienbedingungen aus dem Blickfeld geraten können. So wird vermutet, dass die Anstrengungen von Lehrenden in Ergebnissen von Studentenfüragungen nicht sichtbar sind, da ein vielleicht vorhandener positiver Effekt durch die Verschlechterung der Rahmenbedingungen wieder zunichte gemacht worden sein könnte. In der Tat gibt es Studien, die einen (wenn auch begrenzten) Einfluss bestimmter Rahmenbedingungen auf Ergebnisse von Studentenfüragungen belegen.

Die Soziologieausbildung an der TU Dresden kann als exemplarisch für die Situation gelten: Die in den Anfangsjahren seit Gründung des Studienganges 1993 noch sehr gute Betreuungsrelation, die 1998/99 durchaus noch zu spüren war, hat sich 2001/02 deutlich verschlechtert. Hinzu kommt, dass Sachmittel und Mittel für studentische Hilfskräfte (z.B. zur Lehrveranstaltungsvorbereitung) massiv gekürzt wurden. Parallel hierzu wurden seit Wintersemester 1998/99 an der TU Dresden wie auch am Institut für Soziologie flächendeckend und regelmäßig Lehrveranstaltungsbewertungen durchgeführt und darauf gedrungen, die Ergebnisse noch im sel-

1 Der Verfasser promoviert an der Philosophischen Fakultät der TU Dresden zum Thema „Leistungsbewertung und Leistungsanreize in der Hochschullehre“, leitet die am Institut für Soziologie angesiedelte Projektgruppe Absolventenfüragungen der TU Dresden und übernahm im Rahmen eines Lehrauftrages am Institut für Soziologie die Leitung der Projektseminare „Evaluation von Hochschulen“ (1998/99) und „Qualitätsmanagement im Öffentlichen Sektor“ (2001/02), aus denen dieser Beitrag entstand. Er stellt daher gewissermaßen eine Gemeinschaftsarbeit mit den Seminarteilnehmern dar. Für Fehler übernimmt jedoch der Verfasser allein die Verantwortung.

ben Semester vorzustellen und mit den Studierenden Verbesserungsmöglichkeiten zu besprechen<sup>2</sup>.

## 2. Forschungsfragen

Angesichts der vorliegenden Situation stellte sich zunächst die Frage, inwieweit und wie stark sich Studienbedingungen und Lehrqualität über den Zeitraum von 3 Jahren hinweg gesehen verändert haben. Aus den von einzelnen Fakultäten der TU Dresden (z.B. Fak. Elektrotechnik) über einen solchen längeren Zeitraum vorliegenden Befragungsergebnissen war bekannt, dass neu eingestellte Lehrende nach und nach durchaus an ein gutes Lehrniveau herangeführt werden können. Vergleiche von Lehrbewertungen über kürzere Zeiträume (2-3 Semester) zeigten hingegen kaum deutliche Unterschiede. Im zweiten Schritt sollte daher der Frage nachgegangen werden, ob es trotz sich verschlechternder Bedingungen über einen längeren Zeitraum möglich war, die didaktische Qualität der Lehre wenigstens zu sichern oder eventuell sogar zu verbessern.

Da die für Lehrveranstaltungsbewertungen verwendeten Fragebögen in der Zwischenzeit mehrfach modifiziert wurden und zudem nur sehr wenige Fragestellungen zu den Rahmenbedingungen des Studiums enthielten, wurde eine eigene Befragung zur Situation von Lehre und Studium durchgeführt. Die Schwerpunktsetzung erfolgte hierbei auf den Untersuchungsgegenstand Studienbedingungen. Der Fragebogen enthielt aber auch einige zentrale Fragen zur didaktischen Qualität der Lehre. Außerdem sollte eine Einordnung der Ergebnisse im Vergleich zu Ergebnissen bundesweiter Befragungen erfolgen. Hierfür wurde ein ursprünglich von der HIS GmbH sowie Hornbostel/Daniel verwendeter Fragebogen im Rahmen des Projektseminars an der TU Dresden weiterentwickelt, dessen veränderte Fragestellungen außerdem auf ihre Tauglichkeit überprüft werden sollten. Da allgemeine methodische und konzeptionelle Fragen zu Lehrveranstaltungsbewertungen und zu Studienbedingungen vom Verfasser bereits an anderer Stelle ausführlicher diskutiert wurden (vgl. Krempkow 1999; Krempkow/Heldt 2000b; Krempkow/Winter 2000), soll hier nicht näher darauf eingegangen werden. Für die Diskussion, welche Aspekte „dozentenattribuierbar“ seien, wird auf Marsh (1997) sowie auf Rindermann (2001) verwiesen, für andere Streitpunkte, wie die Berufspraxisorientierung, auf Teichler (2002).

Im nachfolgenden Beitrag soll daher als Grundlage der Analysen zunächst die Durchführung der Befragung dokumentiert und dann die Zuverlässigkeit und Vergleichbarkeit der Ergebnisse diskutiert werden. Davon ausgehend soll geprüft werden, ob Veränderungen evtl. auch auf potentielle Einflussfaktoren zurückgeführt werden könnten. Anschließend sollen die Ergebnisse zur didaktischen Qualität der Lehre eingehender untersucht werden. Im Ausblick sollen schließlich auch Voraussetzungen und Möglichkeiten angesprochen werden, bei künftigen Befragungen von Studierenden bessere Ergebnisse zu erzielen.

2 Zum genauen Ablauf des Lehrevaluationsverfahrens an der TU Dresden vgl. Krempkow/Heldt (2000b).

### 3. Methode, Durchführung und Auswertung der Befragung

#### 3.1 Methode und Durchführung der Befragung

Trotz der Vielzahl der Ranglisten, die bereits auf dem (Leser-)Markt existieren, bedurfte – wenn man auch das Ziel einer Einordnung in bundesweite Ergebnisse verfolgen will – die Auswahl der Methode keiner großen Überlegung: Es gibt nur eine einzige bundesweite Untersuchung zu Studienbedingungen in der Soziologie, die zum Befragungszeitpunkt einem größeren Publikum zugänglich war und ausführlich wissenschaftlich untersucht wurde. Dies ist die Spiegel-Publikation aus dem Jahre 1993. Diese als Rangliste konzipierte Untersuchung wurde speziell für die Soziologie noch einmal ausführlicher in einem 1996 herausgegebenen Band „Soziologielehre in Deutschland“ ausgewertet (vgl. Hornbostel/Daniel 1996). Da die erste Spiegel-Untersuchung (1989/90) ebenfalls sehr ausführlich in soziologischen Veröffentlichungen (z.B. in der Zeitschrift „Soziologie“, Ausgaben 2/1990 und 1/1991) diskutiert wurden, dürften die Fehler und Schwächen der Herangehensweise bei dieser Untersuchung in Fachkreisen hinlänglich bekannt sein. Dennoch soll hier in aller Kürze auf die unserer Ansicht nach wesentlichsten Kritikpunkte eingegangen werden. An erster Stelle wären die Anzahl und die Auswahl der 12 bzw. 18 an jedem Fachbereich zu befragenden Studenten zu nennen. Die Spiegel-Autoren sprechen zwar von einer Zufallsauswahl, davon kann jedoch keine Rede sein. Vielmehr handelt es sich um eine eher willkürliche Auswahl, bei sehr viel gutem Willen um eine Quotenauswahl. (Die Interviewer wurden angewiesen, an mindestens drei verschiedenen Orten – Mensa, Hörsaal, Bibliothek – eine möglichst gleich große Anzahl an Befragten auszuwählen; vgl. Spiegel-special 3/1993.) Bei einer solchen Auswahl der zu Befragenden ist damit zu rechnen, dass mehr oder weniger große systematische Verzerrungen auftreten können.

Zu rechtfertigen ist die Wahl dieses Verfahrens deshalb weder methodisch noch theoretisch, allerdings pragmatisch: Da den Spiegel-Autoren von den meisten Hochschulen die Versendung der Fragebögen mit Unterstützung der Studentensekretariate bzw. Immatrikulationsämter verweigert wurde, hatten sie keine andere Wahl. Sie mussten ihre Interviewer in die Hochschulen schicken, um an die Urteile der Studierenden zu gelangen.

Während das Befragungsverfahren massiv in der Kritik stand, gilt dies kaum für die verwendeten Fragestellungen, die zudem in der Mehrzahl bereits seit mehreren Jahren auch von HIS für Studentebefragungen verwendet wurden. Von der Erweiterung der Frage 7 auf das Grundstudium und einem später noch zu erläuternden Ergänzungsblatt abgesehen, wurde daher derselbe Fragebogen verwendet und mit denselben Vorgaben wie beim Spiegel im Sommersemester 1998 in einer kleinen Vorstudie zunächst 18 Studierende befragt. Abweichend wurden hier für die Bewertung aber auch Studierende des Grundstudiums zugelassen, wodurch aber erstens kein bedeutsamer Einfluss auf das Urteil zu erwarten ist (vgl. Hage 1996; Krempkow 1997, 1999, 2000) und zweitens auch diese nicht gerade kleine Gruppe von Studierenden, die die Wahrnehmung eines Faches an der Hochschule und in der Öffentlichkeit entscheidend mitprägen, in die Untersuchung einfließen konnte.

In der ersten Hauptbefragung im Wintersemester 1998/99 wurden im Gegensatz zum Sommersemester 1998 und zum Spiegel-Ranking 1993 keine face-to-face-

Interviews durchgeführt, sondern eine standardisierte schriftliche Befragung mittels eines maschinenlesbaren Fragebogens. Hierdurch konnte einem der wesentlichsten Kritikpunkte begegnet werden, der zu geringen Anzahl der Befragten beim Spiegel-Ranking 1993. Außerdem sollte überprüft werden, inwieweit man bei einer schriftlichen Befragung mit deutlich mehr Teilnehmern wieder zu den selben Ergebnissen kommen würde wie bei der nach dem Spiegel-Ranking durchgeführten mündlichen Befragung mit 18 Teilnehmern, wobei natürlich auch hier der zeitliche Abstand von einem Semester zu berücksichtigen wäre. Das Fragebogenlayout musste hierfür leicht abgewandelt werden, es wurden jedoch im Wesentlichen die gleichen Fragestellungen verwendet. Die Befragung war wie die für das Spiegel-Ranking als Befragung auf Fachebene konzipiert und wurde im Dezember 1998 und Januar 1999 in vier als weitgehend repräsentativ für das Institut eingeschätzten Vorlesungen durchgeführt.<sup>3</sup> Der größte Anteil entfiel dabei auf die Methoden- und die Theorienvorlesungen. Hierfür soll an dieser Stelle den Lehrenden des Institutes ausdrücklich gedankt werden.

Zusätzlich zu den im Sommersemester 1998 verwendeten Fragestellungen wurde im Wintersemester 1998/99 auch nach angestrebtem Abschluss und Geschlecht der Studierenden gefragt. Studierende des Diplomstudienganges sind mit 49,7 Prozent der Befragten etwas häufiger vertreten als Studierende des Magisterstudienganges mit 42,6 Prozent (Rest: Lehramt). Die Geschlechterverteilung entspricht mit 65 Prozent weiblichen Studierenden in etwa derjenigen in den entsprechenden Statistiken des Immatrikulationsamtes der TU Dresden.

Für die zweite Hauptbefragung im Wintersemester 2001/02 wurde der Fragebogen in einigen Formulierungen nochmals leicht überarbeitet und auf mehrheitlichen Wunsch der Lehrveranstaltungsteilnehmer die Frage (6b) zur Fachstudienberatung zusätzlich aufgenommen. Außerdem wurden nun in der Erläuterung (nach Frage 5) unter Dozenten explizit auch Gastdozenten mit erfasst. In Frage 11 ist statt der in der Soziologie üblichen Bezeichnung „Examen“ nunmehr von „Prüfungen“ die Rede. Bei Frage 12 wird danach gefragt, „Haben Sie das Gefühl, dass ...“, da nicht suggeriert werden sollte, dass Studierende die Orientierung an der Berufspraxis tatsächlich einschätzen können und hier der Charakter der subjektiven Meinung stärker schon in der Fragestellung betont werden sollte (vgl. Lenz/Wagner 2001). In Frage 18 wurde präzisiert, was alles bei der Einschätzung der individuellen Betreuung erfasst werden sollte. Und schließlich wurde den Befragten bereits im Einleitungstext zum Fragebogen der Hinweis gegeben: „Wenn Sie eine Frage nicht beantworten können, lassen Sie sie bitte frei!“ So sollte schon im Ansatz der Eindruck verhindert werden, die Befragten würden zu einer Einschätzung „gezwungen“ die sie sich vielleicht gar nicht zutrauen. Es wurden letztlich jedoch von allen Befragten so gut wie alle Fragen beantwortet.

3 Um eine größere Anzahl von Befragten zu erreichen, ohne daß eine postalische Befragung nach einem Quoten- oder Zufallsauswahlverfahren möglich gewesen war, hätte theoretisch auch eine Vollerhebung vorgenommen werden können. Dies konnte jedoch in der Praxis aufgrund der Auswahl der Lehrveranstaltungen, der Anzahl der Befragten und wegen der daraus resultierenden Semesterverteilung bei den Befragten nicht vollkommen verwirklicht werden. Hierfür wäre auch eine Befragung in Seminaren und Hauptseminaren notwendig gewesen, welche jedoch mit einem vertretbaren Aufwand nicht möglich war.

Beide Fragebogenvarianten (der von 1999 [n=98] und der von 2002 [n= 120] wurden in einer quasiexperimentellen Untersuchungsanordnung nochmals auf ihre Zuverlässigkeit und evtl. aus Frageumformulierungen resultierende signifikante Unterschiede überprüft. Die Analysen ergaben, dass bei keiner der geringfügig umformulierten Fragen signifikante Unterschiede zwischen den beiden eingesetzten Fragebogenvarianten auftraten. Daher wurden die Ergebnisse der beiden Fragebogenvarianten zusammengefasst.

Die Anzahl der Befragten betrug damit insgesamt 427 Studierende (1998/99: 208, 2001/02: 219). Die Durchführung der Befragung wurde als Vollerhebung in ausgewählten Lehrveranstaltungen konzipiert (Vorlesungen Methoden, Theorien), wobei der „Rücklauf“ nach Einschätzung der beauftragten Studierenden über 90% betrug. Die Repräsentativität der Befragung kann für Soziologiestudierende der TU Dresden in Bezug auf die anhand der aktuellsten vorliegenden Lehrberichtsstatistiken (vom Studienjahr 1999/2000) überprüfbar Merkmale Studiengang (Befragte: 62% Diplom, Rest Magister) Geschlecht (72% weiblich) sowie Verteilung nach Grund- und Hauptstudium (56% Grundstudium) mit maximal 5 Prozentpunkten Abweichung als gegeben angesehen werden.

### 3.2 Auswertung der Befragung

Die Auswertung der Befragung wurde mit einer an der TU Dresden im Projekt Evaluation der Lehre vorhandenen Fragebogeneinlesestation durchgeführt. Einlesefehler sind wegen der integrierten Kontrolle der Erkennungszuverlässigkeit und Kontrolle der Daten auf Plausibilität vor Überführung in einen Datensatz so gut wie ausgeschlossen. Lediglich schwache Bleistiftmarkierungen und Markierungen mit einem hellen Faserstift (die bei Beachtung der üblichen Hinweise bei der Befragung nicht auftreten dürften) können vom Programm nicht zuverlässig erkannt werden. Solche Markierungen wären aber bei der ohnehin zur Auswertung der Kommentare notwendigen Durchsicht der Fragebögen aufgefallen und hätten von Hand eingegeben werden können, was aber nicht notwendig war.

Zur Berechnung der Mittelwerte wurden die Fragen 1 (Überfüllung von LV) und 2 (Ausfall von LV) umgepolt, so dass der Wert „1“ analog dem Schulnotensystem in der Bundesrepublik Deutschland generell einem positiven Urteil entspricht, der Wert „6“ einem negativen Urteil. Zur Darstellung der Mittelwertprofile wurden wie auch beim Spiegel-Ranking und der im Sommersemester durchgeführten Befragung nur die Fragen 1 bis 18 verwendet. Bei den Fragen 15 und 16 wurden jedoch anders als beim Spiegel-Ranking die Verfügbarkeit von Arbeitsplätzen in den Bereichen Bibliothek und PC-Pool getrennt erfragt. In den vorhergehenden Befragungen wurde dies über alle Arbeitsbereiche hinweg erhoben, deshalb erscheint im Mittelwertprofil hierfür bei den beiden Bereichen jeweils ein und derselbe Wert.

Frage 19 (Hochschulwahl) ist ein Indikator für die Studierzufriedenheit, der ebenfalls in der Befragung im Sommersemester 1998 und im Spiegel-Ranking verwendet wurde. Hier fällt das Ergebnis im Vergleich zu 1998/99 deutlich negativer aus. Nur noch 23 Prozent der Befragten gaben an, wieder an der TU Dresden studieren zu wollen, wenn für die Wahl der Hochschule allein die Qualität der Ausbildung ausschlaggebend wäre (1998/99: 31%). Das genau umgekehrte Verhältnis er-

gibt sich für die Antwortvorgabe „mein Studium an einer anderen Universität“, wofür sich 2001/02 31% entscheiden würden (1998/99: 23%). Allerdings fühlten sich mit 38 Prozent 2001/02 auch etwas mehr nicht ausreichend informiert (1998/99: 33%). (Wie eingangs bereits erwähnt, lag die letzte Veröffentlichung zu Studienbedingungen und Studienqualität in der Soziologie bereits fast 10 Jahre zurück und der CHE-Studienführer zu Soziologie/Sozialwissenschaften war zum Befragungszeitpunkt noch nicht veröffentlicht.

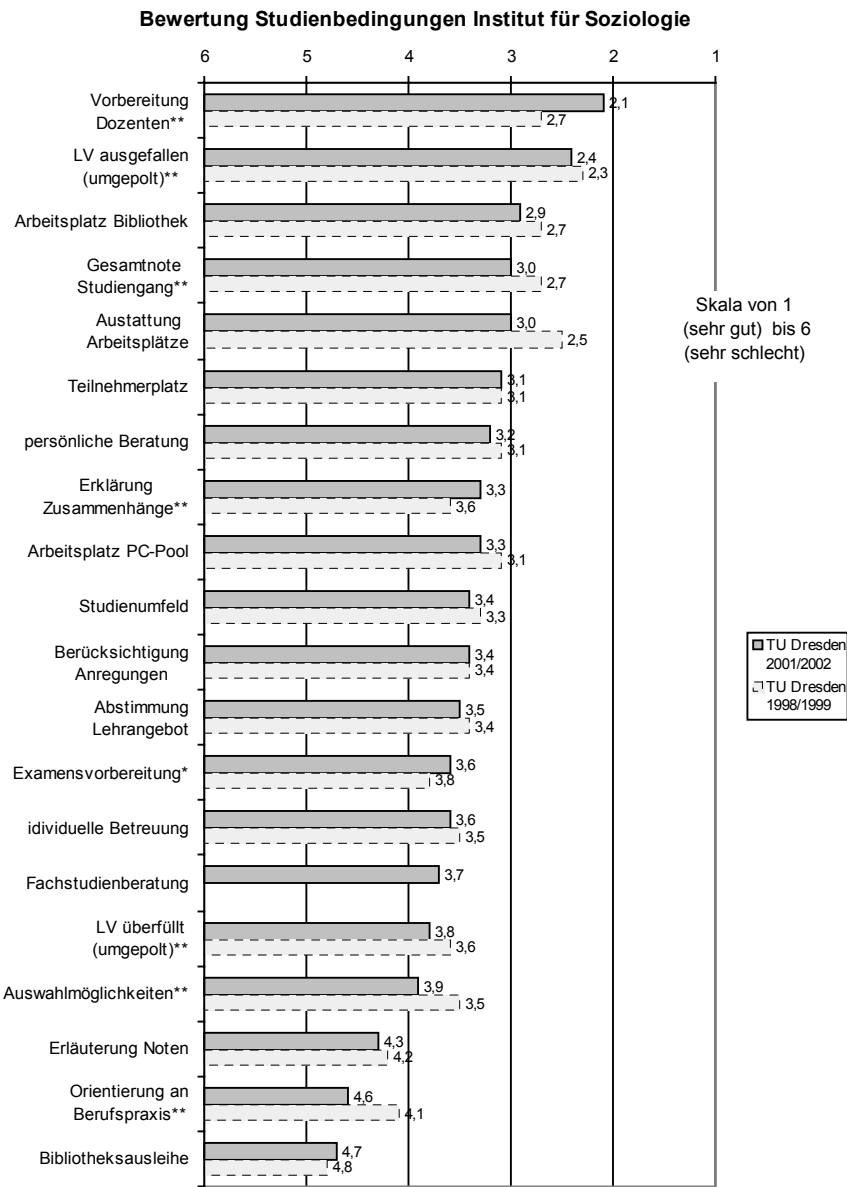
#### 4. Die Ergebnisse der Befragungen 1998/99 und 2001/02 im Überblick

Der Vergleich 2002-1999 zeigt: Die Soziologieausbildung an der TU Dresden schnitt im Urteil ihrer Studierenden im Wintersemester 2001/2002 in vielen der untersuchten Aspekte negativer ab als im WS 1998/99. Dies betrifft die Aspekte Überfüllung und Ausfall von Lehrveranstaltungen, insbesondere aber die Auswahlmöglichkeiten des Lehrangebotes, die Orientierung an der Berufspraxis und die Ausstattung der Arbeitsplätze. Hier fielen die Einschätzungen der Studierenden deutlich und signifikant negativer aus als 1999. Dies gilt auch für die Gesamteinschätzung der Studienbedingungen am Institut für Soziologie.

Signifikant positiver bewertet wurden im Vergleich zum WS 1998/99 dagegen die Vorbereitung der Dozenten (Dozentenengagement), die Erklärung von Zusammenhängen (Didaktik) und die Examens- bzw. Prüfungsvorbereitung. Da – wie bereits erwähnt – seit April 1999 an der TU Dresden flächendeckend studentische Lehrveranstaltungsbewertungen existieren, liegt es nahe, dies hierauf zurückzuführen. Eine zwingende Kausalität kann hieraus natürlich nicht geschlossen werden, andere plausible Erklärungen (z.B. deutlich bessere Lehrqualität durch Wechsel der Dozenten, wobei die neu eingestellten didaktisch deutlich besser sein könnten) sind jedoch auszuschließen. Bei den restlichen der insgesamt 20 erfragten Aspekte konnten keine signifikanten Unterschiede nachgewiesen werden. Allerdings sind bei 16 der untersuchten Studienbedingungen tendenziell negativere Einschätzungen zu beobachten (vgl. folgende Abbildung 1).

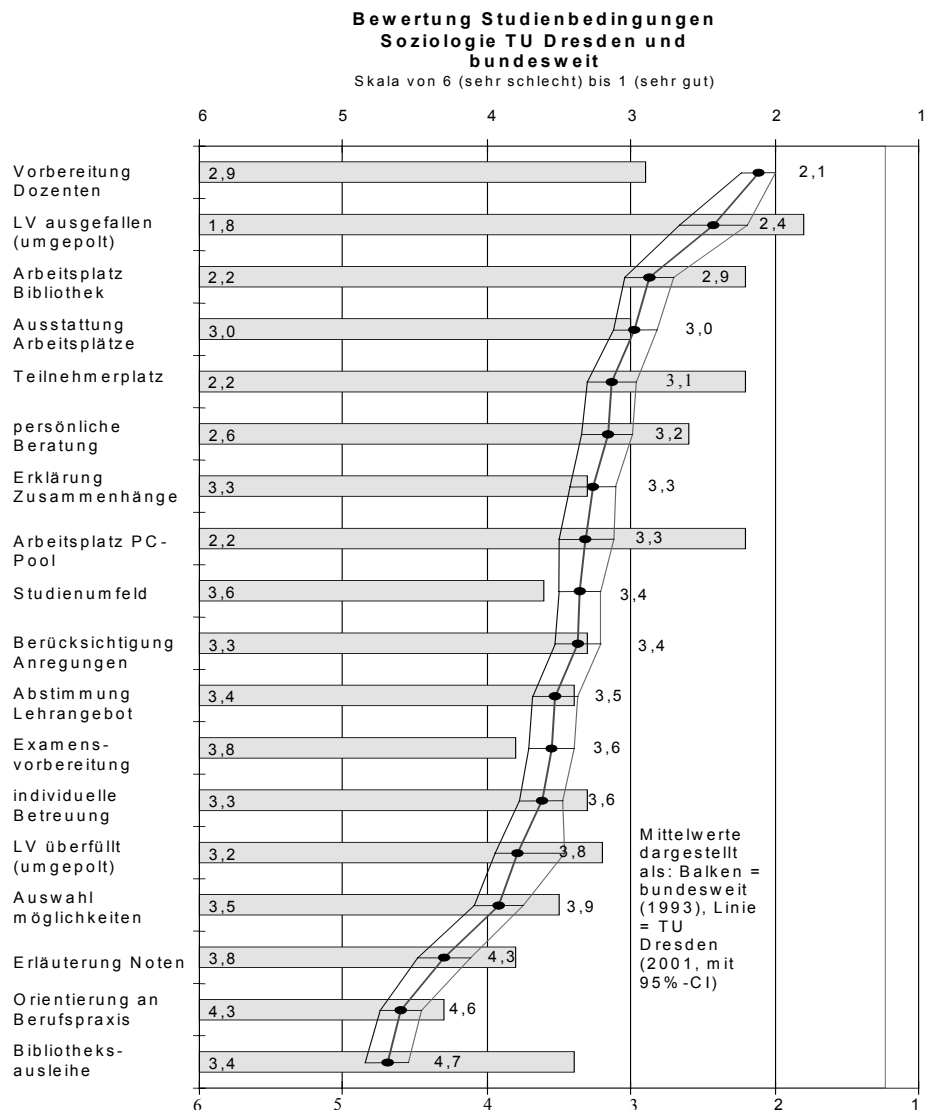
Das wie von Hornbostel/Daniel (1996) aus den einzelnen Fragen rechnerisch ermittelte Gesamturteil fällt 2001/02 mit einem Mittelwert von 3,4 nur geringfügig negativer aus als 1998/99 (3,3). Da dieses Verfahren der rechnerischen Ermittlung eines Gesamturteils – wie noch zu beschreiben sein wird – jedoch umstritten ist, wurde hier wie auch 1999 das subjektive Gesamturteil der Studierenden zum Vergleich herangezogen. Dieses fällt 2001/02 mit einem Mittelwert von 3,0 im Vergleich zu 1998/99 (2,7) deutlich negativer aus. Die wesentlichen Kritikpunkte der Studierenden bleiben trotz Verbesserungen bei Didaktik und Dozentenengagement dieselben und die TU-Soziologie würde, wenn sie an einem Ranking analog dem Spiegel-Ranking teilnähme, auch nach den Ergebnissen der Befragung 2001/02 im Bundesvergleich in vielen Aspekten sehr wahrscheinlich unterdurchschnittlich abschneiden (vgl. Abbildung 2).

Abbildung 1: Vergleich der Ergebnisse der Studentenbefragungen 1998/99 und 2001/02



*Erläuterung:* Die Graphik stellt die Mittelwerte der studentischen Einschätzungen dar, wobei die dunkleren Balken mit durchgezogener Umrahmung die der Befragung 2001/02 bezeichnen. Die helleren Balken mit gestrichelter Umrahmung zeigen die Mittelwerte der Befragung 1998/99. Signifikante Unterschiede sind mit einem Stern\* hinter der Beschriftung des jeweiligen Aspektes gekennzeichnet, sehr signifikante Unterschiede mit\*\*.

Abbildung 2: Einordnung der Ergebnisse der Studentenfragung 2001/02 in bundesweite Ergebnisse



*Erläuterung:* Die Graphik vergleicht die Mittelwerte der studentischen Einschätzung am Institut für Soziologie der TU Dresden mit denen der bundesweiten Befragung von Hornbostel/Daniel (1993). Die durchgezogene Linie mit den zwei „Nebenlinien“ zeigt die Bewertungen an der TU Dresden mit 95%-Confidence-Intervall. Die Balken zeigen die bundesweiten Ergebnisse. Als signifikant sind Unterschiede immer dann einzuschätzen, wenn die Linie inkl. „Nebenlinien“ die Balkenspitze nicht schneiden. So wird die „Vorbereitung der Dozenten“ an der TU signifikant besser bewertet als bundesweit. Bei der „Erklärung von Zusammenhängen“ zeigt sich kein signifikanter Unterschied.



## 5. Diskussion der Zuverlässigkeit und Vergleichbarkeit der Ergebnisse

### 5.1 Das Gesamturteil und seine Zuverlässigkeit

Bekanntlich gibt es neben systematischen Fehlern, die anhand verschiedener potentieller Einflussfaktoren später noch genauer untersucht werden, auch die Möglichkeit zufälliger Fehler. Ein relativ einfaches Verfahren ist die Abschätzung dieses zufälligen Fehlers – unter der theoretischen Annahme einer Zufallsauswahl – mittels Berechnung des Konfidenzintervalls.

Da Hornbostel/Daniel (1996) die Vertrauensbereiche jedoch nur für die Gesamtränge auswiesen, ist hier aus Gründen der methodischen Vergleichbarkeit nun die Bildung eines Gesamttranges notwendig. Hornbostel/Daniel haben den Gesamttrang berechnet, indem sie einfach die Mittelwerte der einzelnen Variablen aufsummierten und durch die Anzahl der Variablen teilten. Dies wurde mehrfach als „Äpfel-und-Birnen-Vergleich“ kritisiert. Besser geeignet wären wohl eine Berechnung der Vertrauensbereiche und deren Ausweisung für jedes einzelne Kriterium. Ohne Zugriff auf den Datensatz oder zumindest die Ausweisung von Streuungswerten ist dies aber nicht möglich. Bei entsprechender Vorsicht und der Berücksichtigung der bereits erwähnten systematischen Fehler lässt sich, wenn man zunächst allein den zufälligen Fehler in Betracht zieht, aber zumindest die Größe des Vertrauensbereichs in der Stichprobe der TU Dresden abschätzen. Für den Mittelwert von 3,39 ergibt sich dann ein Vertrauensbereich von 3,32 bis 3,46. Dieser überlappt sich lediglich mit den in der Spiegel-Gesamtrangliste ausgewiesenen Vertrauensbereichen der am negativsten bewerteten Universitäten. Rein mathematisch-statistisch besehen, wäre die TU Dresden also unter den letzten Universitäten gelandet, wenn man sie genau wie dort geschehen bewertet hätte. Zwischen den Datenerhebungen 1993 und 1998 bzw. 2001 liegen aber einige Jahre, und wenn man die fast überall stattfindenden Mittelkürzungen im Hochschulbereich in Betracht zieht, dürfte die Situation an den Hochschulen sich insgesamt eher verschlechtern haben, so dass sich evtl. auch an den dort genannten Universitäten heute schlechtere Bewertungen ergeben würden.

Auch das Vorhandensein systematischer Fehler spräche dagegen, allein der „reinen Statistik“ zu vertrauen. Und schließlich muss in Betracht gezogen werden, dass es sich bei der TU Dresden um eine Ost-Universität handelt, für die im Spiegel-Ranking 1993 aufgrund der spezifischen Situation kein Fachbereichs- sondern nur ein Hochschulvergleich vorgenommen wurde. Inzwischen gibt es jedoch Erfahrungswerte mit Rankinglisten. Und diese Erfahrungen besagen, dass kaum jemals – selbst bei Anwendung verschiedener Erhebungsmethoden (vgl. Hornbostel/Daniel 1996: 22; Hornbostel 1999: 86; Hornbostel 2001) – eine Universität bei dem einen Ranking einen Spitzenplatz erzielte, und bei einem anderen Ranking in der Schlussgruppe landete. Verschiebungen ergaben sich fast immer nur innerhalb der einzelnen Gruppierungen bzw. zwischen Spitzengruppe und Mittelfeld, oder zwischen Mittelfeld und Schlussgruppe. Selbst bei größter Vorsicht in der Interpretation muss man also selbstkritisch einräumen, dass die Soziologie der TU Dresden mit hoher Wahrscheinlichkeit wohl auch heute bei einem Ranking unterdurchschnittlich abschneiden würde. Bestenfalls wäre sie Mittelmaß.

Neben der Aufsummierung der einzelnen Variablen wie bei Hornbostel/Daniel, oder der Gewichtung der einzelnen Fragen wie im Focus-Ranking 1997 (wobei die Festlegung der Gewichtungsfaktoren leider nicht begründet wurde) gibt es noch eine weitere in den Sozialwissenschaften verbreitete Möglichkeit, eine Gesamteinschätzung zu erhalten: indem man die Betroffenen am Ende des Fragebogens selbst zu Wort kommen lässt. Hierbei können sie ihre ganz persönliche subjektive Gewichtung vornehmen, die man dann dem rechnerisch ermittelten Gesamturteil gegenüberstellen könnte. Dies wurde hier getan, und das Ergebnis sieht mit einem Mittelwert von 3,0 schon etwas positiver aus. Allerdings wäre die TU Dresden selbst bei Einordnung des subjektiven Gesamturteils (das bei anderen Hochschulen wahrscheinlich ebenfalls deutlich positiver ausfallen würde) in das rechnerisch ermittelte Gesamturteil immer noch Mittelmaß.

## 5.2 Vergleichbarkeit der Gesamturteile

Neben den bereits genannten Quellen wird häufig das Vorhandensein potentieller Einflussfaktoren genannt, die zu einer systematischen „Verzerrung“ der studentischen Urteile geführt haben können. Solche Einflussfaktoren sind nach inzwischen vorliegenden empirischen Studien für Deutschland und den einschlägigen amerikanischen Studien (insbesondere Herbert W. Marsh 1982, 1997) das Studieninteresse, die Leistungsfähigkeit und -bereitschaft der Studierenden, die Überfüllung der Kurse, sowie weitere Rahmenbedingungen.

Der stärkste Zusammenhang wurde bei der Untersuchung des Studieninteresses gefunden, und zwar sowohl auf Fachbereichsebene (vgl. Hage 1996; Krempkow/Heldt 2000a), als auch auf Ebene einzelner Lehrveranstaltungen (vgl. Kromrey 1996). Jedoch hegen einige Autoren Zweifel daran, ob das Studieninteresse als „echte“ Biasvariable zu bezeichnen wäre (vgl. Spiel/Gössler 2000; Daniel 1998). Andere Autoren führen zudem aus, dass einerseits durch eine interessante Gestaltung Interesse geweckt werden kann und andererseits durch eine adäquate Information der Studienanfänger und -Interessenten „falsche“ Erwartungen bereits im Ansatz vermieden werden könnten (vgl. Schwarzkopf 2001). Einig sind sich die Autoren aber darin, dass das Interesse der Studierenden in jedem Fall erfasst und bei stärkerem Einfluss in die Ergebnisinterpretation einbezogen werden sollte.

Der am häufigsten untersuchte Einflussfaktor ist wohl der Einfluss der Noten bzw. der Notenerwartung auf die Ergebnisse von Studentenforschungen zur Lehrqualität. Um eine „Verzerrung“ und Beeinträchtigung der Vergleichbarkeit festzustellen, muss dies jedoch über mehrere Lehrveranstaltungen hinweg betrachtet werden. Sowohl ältere Untersuchungen (Übersicht siehe Krempkow 1999) als auch inzwischen erfolgte neuere Untersuchungen hierzu konnten jedoch entweder keine (Daniel 1998; Schwarzkopf 2001) oder nur relativ kleine Zusammenhänge feststellen (Krempkow/Heldt 2000c). Gleiches gilt für die Leistungsbereitschaft und auch für die Kursgröße, die als Rahmenbedingung der Lehrveranstaltungen einbezogen wurde.

Schließlich wurde in mehreren Untersuchungen noch der Einfluss von Schwierigkeitsgrad bzw. Anspruchsniveau der Lehrveranstaltung untersucht. Beispielhaft soll hier eine Untersuchung erwähnt werden, die von einem verzerrenden Einfluss

ausgeht (Esser 1997), und eine, die diesen Einfluss relativiert (Krempkow/Heldt 2000). Oft wurde der Schwierigkeitsgrad von Lehrveranstaltungen wie bei Esser mittels Skalen erfragt, die nur eine positive (höhere Anforderungen) und eine negative Bewertung (niedrigere Anforderungen) ermöglichen. Hier lassen sich dann z.T. die erwarteten Zusammenhänge finden und als Verzerrung der Ergebnisse interpretieren.

Untersucht man den Einfluss der Leistungsanforderungen jedoch auf einer Skala, die zusätzlich zu hohen und niedrigen Anforderungen auch die Antwortvorgabe „optimal“ (z.B. in der Mitte der Skala) enthält, so zeigt sich ein anderes Bild. Gegenübergestellt werden kann dann jeweils in Gruppen die Einschätzung der Lehrqualität von Studenten, die die Anforderungen als (zu) hoch, optimal, oder (zu) niedrig empfinden. Zwar werden auch hier Aufbau und Struktur, Interessenförderung und in geringem Maße die Didaktik der Lehrveranstaltung bei als „zu hoch“ empfundenen Anforderungen etwas schlechter (0,2 Skalenpunkte, Skala 1-5) bewertet als bei hohen Anforderungen. Deutlich schlechtere Bewertungen (0,6-0,8 Skalenpunkte) erhalten die Lehrveranstaltungen jedoch bei als „zu niedrig“ empfundenen Anforderungen. Die besten Bewertungen erhalten die Lehrveranstaltungen aber immer dann, wenn die Anforderungen als „optimal“ empfunden werden. Dieser Effekt wurde bereits 1999 festgestellt (vgl. Krempkow/Heldt 2000a, 2000d sowie Krempkow/Winter 2000) und hat sich inzwischen mehrfach bestätigt. Das Ergebnis deckt sich mit Erfahrungen von Didaktikern und Lernpsychologen: Optimale Lehr- und Lernerfolge können demnach nur erzielt werden, wenn auch die Anforderungen auf die Lernenden abgestimmt sind.

### 5.3 Einflussfaktoren auf das Gesamturteil

Die erst in den letzten Jahren intensiver untersuchten Einflussfaktoren konnten Hornbostel/Daniel z.T. noch nicht berücksichtigen, in unseren Untersuchungen wurden jedoch die relevanten potentiellen Einflussfaktoren untersucht. Das Ergebnis war aber größtenteils ebenso ernüchternd wie die Ergebnisse zuvor: Unter Einbeziehung der potentiellen Einflussfaktoren und der an anderen Hochschulen festgestellten Zusammenhänge solcher Variablen, die nichts mit der Studienqualität zu tun haben, müssten die Ergebnisse zur Dresdner Soziologie eher nach unten statt nach oben korrigiert werden.

So ist z.B. das Fachinteresse, das vom Spiegel (1993) nicht erfragt wurde, gemessen an der „Wichtigkeit des speziellen Fachinteresses bei der Studienfachwahl“ bei den Soziologiestudierenden an der TU Dresden (Mittelwert 2,0) wie auch schon 1999 stärker ausgeprägt als im Bundesschnitt (2,2). In einer für die Bundesrepublik repräsentativen Untersuchung der Universität Konstanz (vgl. Hage 1996) wurde, ebenso wie bei vorhergehenden anderen Untersuchungen ein *positiver* Zusammenhang des Studieninteresses mit der subjektiven Gesamtbewertung der Studiensituation festgestellt (Rangkorrelationskoeffizient  $r=0,35$ , Signifikanz (einseitig)  $p=0,15$ ). Das hieße, wenn das Studieninteresse der Dresdner Soziologiestudenten nicht überdurchschnittlich hoch wäre, hätte die TU wahrscheinlich noch schlechter abgeschnitten. Für die 2001/02 durchgeführte Befragung konnte zwar kein nennenswerter signifikanter derartiger Zusammenhang festgestellt werden ( $r=0,06$ , Si-

gnifikanz  $p=0,19$ ). Dies schließt aber sein Vorhandensein nicht mit letzter Sicherheit aus.

Ein etwas größerer positiver Zusammenhang des subjektiven Gesamturteils lässt sich mit der erwarteten Abschlussnote feststellen, die als Maß zwar nicht für die tatsächliche Abschlussnote, wohl aber für die Selbsteinschätzung der Studenten und deren Einfluss auf das Gesamturteil dienen kann ( $r=0,12$   $p=0,053$ ). Dies gilt aber ebenfalls nur für das subjektive Gesamturteil, beim rechnerisch ermittelten Gesamturteil ist der Zusammenhang nur noch halb so groß und auch die „Irrtumswahrscheinlichkeit“ steigt rapide ( $r=0,07$ ,  $p=0,18$ ). Man darf die Größe dieses Einflusses also nicht überbewerten. Für die Interpretation, ob nun die TU Dresden sich bei einer besseren Selbsteinschätzung der Studenten (sie liegt mit einer durchschnittlichen erwarteten Abschlussnote von 2,6 etwa in der Skalenmitte) fehlt jedoch der Bundesdurchschnitt. So kann man nur vermuten, dass dies der Fall sein könnte, wenn auch in nur sehr geringem Ausmaß. Die tatsächliche durchschnittliche Abschlussnote lag (mit Abweichungen von +/- 2 Zehntelnoten wie in den vorhergehenden Studienjahren) bei 1,9 (Diplom + Magister-HF,  $n=11$  Absolventen), die der Zwischenprüfungen bei 2,5 (Diplom + Magister-HF,  $n=60$ ) – vgl. Lehrbericht TU Dresden 1999/00: 226 ff.

Ein dritter, potentiell verzerrender Einflussfaktor kann der Grad der Überfüllung sein. Hier muss man aber bereits davon ausgehen, dass er direkt etwas mit der Studienqualität zu tun hat. Man muss auch in Rechnung stellen, dass es nicht unbedingt im Verantwortungsbereich der Lehrenden des Institutes liegt, wenn aufgrund der von den Studierenden empfundenen Überfüllung der Fachbereich insgesamt schlechter beurteilt wird. Die Autoren des Spiegel-Rankings jedenfalls haben die Frage danach, ob die wichtigen Lehrveranstaltungen eher „sehr selten“ oder eher „sehr häufig“ überfüllt sind, und ob man für die Pflichtveranstaltungen „jeweils genau in dem Semester einen Teilnehmerplatz bekommt, in dem es nach der Studienordnung am günstigsten wäre“, in die Bewertung der Studienqualität einbezogen. Außerdem wurden Fragen zur Einschätzung der Beratung und Betreuung aufgenommen. Bei der Auswertung stellte sich heraus, dass die wichtigen Lehrveranstaltungen an der TU Dresden mit einem Mittelwert von 3,8 häufiger als überfüllt eingeschätzt wurden als im Bundesschnitt (3,2), und dass man auch seltener einen Teilnehmerplatz in Pflichtveranstaltungen erhält (3,1) als im Bundesschnitt (2,2). Ein Zusammenhang dieser Variablen mit der subjektiven Gesamtbewertung ist hier bereits nachweisbar, aber nur schwach ausgeprägt ( $r=0,2$   $p=0,003$ ).

Die „individuelle Betreuung der Studierenden wurde wiederum negativer bewertet (3,6) als im Bundesschnitt (3,3). Etwas stärker und sehr signifikant wäre lediglich der Zusammenhang von individueller Betreuung und subjektiver Gesamtbewertung ( $r=0,44$   $p=0,000$ ). Aber um dies als verzerrende Biasvariable herauszurechnen, die nichts mit der Studienqualität zu tun hat, müsste man die Hochschulen als reine Wissenszulieferbetriebe auffassen, die dann mit der Betreuung von Studierenden nichts zu tun haben dürften.

Als weitere potentiell verzerrende Einflussfaktoren sollen nun noch die Bewertungen einiger Rahmenbedingungen des Studiums wie das Studiumumfeld und die Bibliothekssituation untersucht werden. Das Studiumumfeld wurde bewertet, indem danach gefragt wurde, ob „der Unterrichts- und Aufenthaltsbereich der Institute und Seminare so gestaltet“ ist, dass die Studierenden dort „sehr gern“ oder „nur

ungern“ arbeiten. Diese Frage war 1998/99 insbesondere wegen dem zu diesem Zeitpunkt erst kürzlich erfolgten Umzug des Institutes an die Peripherie des Campus interessant. Im Sommersemester 1998 wurde dies in der Tat deutlich schlechter (4,3) als im Bundesschnitt (3,6) beurteilt und hatte einen starken Zusammenhang mit dem Gesamturteil. Im Wintersemester 1998/99 (Mittelwert 3,3) wie auch 2001/02 (Mittelwert 3,4) fiel die Bewertung des Studienumfeldes deutlich besser aus und lag damit über dem Bundesschnitt. Der Zusammenhang von  $r=0,4$  mit  $p=0,00$  kann sich also nur positiv auf das Gesamturteil ausgewirkt haben und bei Berücksichtigung des „verzerrenden“ Einflusses würde das Gesamturteil schlechter ausfallen.

Als weiterer potentieller Einflussfaktor käme nun noch die Bibliothekssituation in Betracht. Die Ausleihmöglichkeiten in der Bibliothek wurden mit einem Mittelwert von 4,7 deutlich negativer bewertet als im Bundesschnitt (3,4). Dies bleibt nicht ohne Einfluss auf das rechnerisch ermittelte Gesamturteil, der Einfluss ist ( $r=0,2$ ,  $p=0,03$ ) aber relativ schwach.

Am stärksten und in allen nachfolgend genannten Fällen sehr signifikant ( $p=0,000$ ) hängt das subjektive Gesamturteil hingegen mit Variablen zusammen, die als direkter Bestandteil der Studienqualität aufgefasst werden können und wohl auch mehrheitlich so aufgefasst werden. Dies ist vor allem die individuelle Betreuung ( $r=0,44$ ). Etwas schwächere, aber immer noch relativ hohe Zusammenhänge finden sich auch mit den Auswahlmöglichkeiten des Lehrangebotes ( $r=0,36$ ), der Orientierung an der Berufspraxis (0,34) sowie mit der Abstimmung von Lehrangebot und Prüfungsanforderungen ( $r=0,33$ ). Beim rechnerisch ermittelten Gesamturteil fallen die Zusammenhänge – mit Ausnahme potentiell verzerrender Biasvariablen wie Fachinteresse ( $r=0,06$ ,  $p=0,37$ ) und Notenerwartung ( $r=0,12$ ,  $p=0,11$ ) – noch deutlich größer aus, wobei letzteres natürlich auch aus der Konstruktion des rechnerisch ermittelten Gesamturteils als Mittel der Variablen resultiert.

Zusätzlich wurde zur Überprüfung der Dimensionalität noch eine Faktorenanalyse durchgeführt (wegen besserer Eignung des größeren Datensatzes und höherer Zuverlässigkeit für die Befragungen 1999 und 2002 als gemeinsame Analyse – vgl. Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin). Hierbei zeigt sich, dass innerhalb der Dimensionen immer nur solche Variablen hoch „laden“, die auch inhaltlich in etwa dieselben Bereiche bezeichnen (Hauptkomponentenanalyse mit Varimax-Rotation). Faktor 1 (bzw. die 1. Komponente), der allein 13,3% der Varianz in der Stichprobe erklärt, bezeichnet Dozentenengagement und Didaktik, da hierauf die entsprechenden Variablen hoch laden (Vorbereitung der Dozenten, die Berücksichtigung von Anregungen und die Erklärung von Zusammenhängen). Faktor 2 bezeichnet v.a. Arbeitsplätze und Umfeld, Faktor 3 Möglichkeiten zur Wissensüberprüfung, Faktor 4 Lehrangebot und subjektives Gesamturteil, Faktor 5 Überfüllung, Faktor 6 Betreuung und Faktor 7 Möglichkeit zum zügigen Studienabschluss (durch Teilnehmerplatz in Pflichtveranstaltungen „in dem Semester, in dem es nach der Studienordnung am günstigsten wäre“) sowie die Berufsorientierung. Dieses 7-Faktoren-Modell<sup>4</sup> erklärt 71% der Gesamtvarianz.

4 Die 7-Faktoren-Lösung erwies sich im Vergleich zu 5- u. 6-Faktorenlösungen mit und ohne Fachstudienberatung als am besten interpretierbar, da hier die Faktoren trennscharf zugeordnet werden konnten. Die Einschätzung der Fachstudienberatung erfolgte nur 2002, daher ist der Wert dieser Faktorladung nur mit Einschränkungen interpretierbar.

Ergänzend soll an dieser Stelle erwähnt werden, dass das CHE (1998, 2001) in einem (Studien-)Entscheidungsmodell für eine Hochschule in ähnlicher Weise die je nach Fachrichtung 30 bis 40 Indikatoren (statistische Daten bzw. Fragestellungen) thematisch zu 9 so genannten „Bausteinen“ zusammenfasste. Solche „Bausteine“ waren u.a. das subjektive Gesamturteil, Studium und Lehre/Lehrangebot/Fachberatung und Betreuung, Ausstattung, Studienergebnis, Studienort und Hochschule/Studierendenstruktur, internationale Ausrichtung und Forschungsaktivitäten. Zum Baustein Studienergebnis wurden beispielsweise zusammengefasst: Notenschnitt, Erfolgsquote bei Prüfungen, ggf. Freiversuche, Studiendauer und Anzahl der Absolventen (aber noch ohne Relation zur Studienanfängerzahl z.B. zum Zeitpunkt des Studienbeginns der jeweiligen Abschlussjahre). Das subjektive Gesamturteil der Studierenden wurde wie in unserer Befragung anhand einer Notenskala (von 1 bis 6) erfasst.

Da die Zusammenhänge des Gesamturteils mit Variablen, die direkt mit der Studienqualität zusammenhängen (sollen), größer sind als mit solchen Variablen, die nicht unbedingt oder nur indirekt etwas mit der Studienqualität zu tun haben, unterstützen auch die Ergebnisse der weiterführenden Analysen die Zuverlässigkeit der Befragung.

#### 5.4 Aussagekraft der Ergebnisse zur didaktischen Qualität der Lehre

Da die didaktische Qualität der Lehre besser, ihre Rahmenbedingungen jedoch schlechter bewertet wurden, sollen die Ergebnisse zur Didaktik noch einmal genauer untersucht werden. Zunächst soll diskutiert werden, ob die Mittelwertunterschiede der Didaktikbewertungen überhaupt bedeutsame Effekte sind. Anschließend soll geprüft werden, ob Veränderungen evtl. auch auf potentiell „verzerrende“ Einflussfaktoren zurückgeführt werden könnten.

Für die Bestimmung, ob Mittelwertunterschiede als bedeutsame Effekte einzuschätzen sind, ist nach Bortz/Döring (2002: 603 ff.) zusätzlich zu Signifikanzanalysen auch die Streuung der Stichproben heranzuziehen (ex-post-Bestimmung der Effektgröße). Anders als Signifikanzanalysen ermöglicht die Effektgrößenbestimmung mittels Effektstärken auch den Vergleich zwischen verschiedenen Studien, da sie anders als Signifikanzanalysen stichprobenunabhängig sind (vgl. z.B. Rindermann/Kohler 2003: 77). In der Untersuchungsplanung wurde von kleinen bis mittleren Effektgrößen ausgegangen. Für die Analyse wurden die Stichproben als unabhängig aufgefasst, da davon auszugehen war, dass nach drei Jahren kaum noch dieselben Studierenden in Grundlagenveranstaltungen anzutreffen sind. Tatsächlich zeigt sich als Ergebnis der Analyse für die Dozentenvorbereitung mit  $d=0,54$  eine mittlere und für die Erklärung von Zusammenhängen mit  $d=0,32$  eine kleine Effektgröße (Klassifizierung nach Bortz/Döring 2002: 604). Hierbei konnte von einer in beiden Stichproben übereinstimmenden Streuung ausgegangen werden. Zwar ist diese Annahme für die Dozentenvorbereitung nicht ganz zutreffend, denn hier unterscheidet sich die Streuung der beiden Stichproben geringfügig, führt man jedoch zusätzlich eine getrennte Streuungsschätzung durch, so erhöht sich die Effektgröße für die Dozentenvorbereitung nur wenig auf  $d=0,56$ . Es kann also auch nach dieser zusätzlichen Untersuchung von einer deutlichen Verbesserung der Einschätzung der Dozentenvorbereitung gesprochen werden. Die Verbesserung der Erklärung von Zusammen-

hängen fällt allerdings geringer aus. Nach den Erfahrungen, die von verschiedenen Projekten zur Evaluation der Lehre mit breiter angelegten Instrumenten zur Einschätzung der Lehrveranstaltungsqualität berichtet wurden, erscheint dies plausibel: Regelmäßige Bewertungen der Lehre vermögen zwar ggf. die Aufmerksamkeit für die Wahrnehmung der Lehraufgaben (wie z.B. zur Vorbereitung der Lehrveranstaltungen) zu steigern. Ohne didaktische Fortbildungsmaßnahmen ist jedoch nur in geringerem Maße die didaktische Qualität der Lehrveranstaltungen zu verbessern. Leider war eine Untersuchung der didaktischen Qualität anhand der Ergebnisse der Lehrveranstaltungsbeurteilung über einen ähnlich langen Zeitraum an der TU Dresden nicht möglich. Eine solche Untersuchung hätte weitere erhellende Analysen erlaubt.

Aber auch wenn sich bedeutsame Effekte zeigen, könnte es sein, dass diese auf andere Einflüsse zurückzuführen sind. Daher soll hier eine Prüfung erfolgen, ob Veränderungen in der Bewertung didaktischer Aspekte (auch) auf potentiell „verzerrende“ Einflussfaktoren zurückgeführt werden müssen. Dies kann in zwei Schritten geschehen: Im ersten Schritt werden die Variablen zu potentiellen Einflussfaktoren auf signifikante Unterschiede zwischen den beiden Stichproben untersucht. Erst wenn sich hier tatsächlich Unterschiede zeigen, erscheint im zweiten Schritt eine ausführlichere Analyse von Zusammenhängen zur didaktischen Bewertung Erfolg versprechend, wie sie für die Gesamturteile vorgenommen wurden. Anderenfalls kann davon ausgegangen werden, dass beide Stichproben in etwa derselben Weise „verzerrt“ bzw. „unverzerrt“ sind.

Potentielle Einflussfaktoren wie Studieninteresse und Notenerwartung, aber auch solche eher soziodemographischen Variablen wie Geschlecht, Fachsemester und angestrebter Abschluss, die die Ergebnisse beeinflussen könnten, unterscheiden sich jedoch nicht signifikant. Daher kann davon ausgegangen werden, dass die Verbesserung der Bewertung didaktischer Aspekte nicht auf z.B. eine Änderung der Notenvergabepraxis oder eine Änderung der Zusammensetzung der Studierenden zurückzuführen ist.

Auch wenn hier sicher nicht mit letzter Sicherheit ausgeschlossen werden kann, dass es evtl. weitere, nicht entdeckte Zusammenhänge gibt, so lässt sich doch die Verbesserung der Einschätzung didaktischer Aspekte zumindest mit großer Wahrscheinlichkeit als eine tatsächlich bessere Bewertung durch die Studierenden einordnen.

## 5.5 Fazit zur Aussagekraft der Ergebnisse

Ein Hauptkritikpunkt an der 1993 von Hornbostel und Daniel durchgeführten Untersuchung, die zu geringe Anzahl der Befragten pro Fachbereich, dürfte bei den in den Wintersemestern 1998/99 und 2001/02 durchgeführten Befragungen ausgemittelt worden sein. Mit 208 bzw. 219 Befragten existiert eine deutlich breitere Datengrundlage als bei den Befragungen von 1993 und vom Sommersemester 1998 mit 18 Befragten. Die Anzahl von insgesamt 427 Befragten liegt sogar leicht über dem optimalen Stichprobenumfang für die Analyse von Mittelwertdifferenzen (vgl. Bortz/Döring 2002: 613). Zwar konnte auch hier ein Zufallsauswahlverfahren bzw. eine Vollerhebung nicht vollständig realisiert werden. Dennoch kann argumentiert

werden, dass der insgesamt etwas zu geringe Anteil von Studierenden im Hauptstudium an den Befragten nach vorliegenden empirischen Studien keinen bedeutsamen Einfluss auf die Ergebnisse hat und im WS 2001/02 alle überprüfbareren Merkmale der Befragten mit den aktuellsten veröffentlichten Statistiken der TU Dresden übereinstimmen. Von der Umformulierung einiger Fragestellungen abgesehen, auf die bereits eingegangen wurde und die keinen statistisch belegbaren Einfluss haben, wären demzufolge keine Argumente gegen einen Vergleich der Ergebnisse einzuwenden. Es fällt jedoch auf, dass die Ergebnisse von den Grundtendenzen her (was insgesamt positiv und was insgesamt negativ bewertet wird) in relativ hohem Maß übereinstimmen. Sowohl wesentliche Kritikpunkte wie die Bibliothekssituation, als auch ihre Einflüsse auf das Gesamturteil sind weitgehend dieselben. Beispielsweise das Studienumfeld wird nicht mehr so negativ wahrgenommen wie noch in der kleinen Vorstudie vom Sommersemester 1998, behält aber im WS 2001/02 wie im WS 1998/99 einen relativ großen Einfluss auf das Gesamturteil. Hierbei „verzerrt“ dieser Einfluss das Gesamturteil in positivem Sinne, während die negativ bewertete Bibliothekssituation das Gesamturteil negativ beeinflusst. Die Einflüsse der negativ und der positiv bewerteten Aspekte beeinflussen das Gesamtergebnis also mit gegenläufiger Tendenz, wobei sich die Einflüsse teilweise wieder aufheben bzw. der „positive“ Einfluss überwiegt. Da für andere potentielle Einflussfaktoren wie Fachinteresse und Notengebung keine oder nur schwache und nicht signifikante Zusammenhänge gefunden werden konnten, kann man davon sprechen, dass das Gesamtergebnis weitgehend frei ist von verzerrenden Einflussfaktoren.

Aufgrund der sehr viel breiteren Datengrundlage und der weitgehenden Freiheit von verzerrenden Einflussfaktoren der Ergebnisse können die Befragungen im Wintersemester 1998/99 und 2001/02 als deutlich aussagekräftiger eingeschätzt werden, als dies noch bei der Befragung im Sommersemester 1998 und erst recht bei der Befragung für das Spiegel-Ranking 1993 der Fall war. Sie eignet sich damit unseres Erachtens nach als ausreichend verlässliche Datengrundlage zur Diskussion der Ergebnisse als Grundlage für Verbesserungsmaßnahmen in den Gremien der Hochschule.

Als zentrale Ergebnisse lassen sich daher formulieren: Die Rahmenbedingungen von Lehre und Studium haben sich aus Sicht der Studierenden in vielen Einzelaspekten und in der Gesamteinschätzung deutlich verschlechtert, wobei eine Verzerrung durch Einflussfaktoren, die nichts mit der Studienqualität zu tun haben, weitgehend ausgeschlossen werden kann. Würden potentielle Einflussfaktoren berücksichtigt, fiel die Verschlechterung z.T. noch deutlicher aus. Dennoch hat sich die Bewertung von didaktischen Aspekten der Lehre seit der Einführung regelmäßiger und flächendeckender studentischer Lehrbewertungen über den langen Zeitraum von 3 Jahren deutlich verbessert. Da sich auch die beiden Teilstichproben bezüglich potentieller Einflussfaktoren nicht signifikant unterscheiden, kann hier eine Verzerrung ebenfalls weitgehend ausgeschlossen werden.

Zwar bleibt als Forschungsdesiderat zu konstatieren, dass ein Langzeitvergleich von Ergebnissen studentischer Lehrbewertungen mit einem breiteren Fragenkatalog zur didaktischen Qualität der Lehre über drei Jahre und länger – dazu möglichst auf Lehrveranstaltungsebene – weitere Erkenntnisse erbringen würde. Aufgrund der Tatsache, dass Fragebögen mit dem Wechsel verantwortlicher Personen häufig erheblich modifiziert werden und dabei z.T. „das Rad neu erfunden wird“ und auf-



grund von Vorbehalten vieler Lehrender gegenüber Auswertungen auf Lehrveranstaltungsebene scheint dies aber nur selten möglich.

## 6. Ausblick

Die Chancen für das untersuchte Institut, im Urteil der Studierenden in Zukunft positiver abzuschneiden, stehen insgesamt nicht so schlecht, wie aufgrund der vorgestellten Ergebnisse vielleicht zunächst angenommen werden könnte. Das Institut hat neben den genannten kritisch eingeschätzten Aspekten auch Potentiale, die sich ausbauen lassen und einige gute Voraussetzungen, die sich aufgrund der analog dem Spiegel-Ranking in dieser Befragung verwendeten Kriterien nicht niederschlagen konnten. Darüber sollte die insgesamt unterdurchschnittliche Bewertung nicht hinwegtäuschen. So ist die Betreuungsrelation, die nach den Angaben in den Lehrberichten berechnet wurde, mit etwa 14 Studienanfängern und ca. 100 Studierenden je Professor im Bundesvergleich noch relativ gut (vgl. Hornbostel/Daniel 1996: 18), wenngleich bei solchen Kennzahlen immer auch noch eine gewisse Skepsis angebracht ist (vgl. Krempkow/König/Winter 2001). Allerdings ist der „Schwund“ von 83% (bzw. die Absolventenquote von 17% – vgl. Krempkow/König/Winter 2001) bis zum 10. Fachsemester im Bundesvergleich (67% – vgl. Hornbostel/Daniel 1996) überdurchschnittlich hoch. Hierbei ist aber zu beachten, dass dies mit vielen Faktoren zusammenhängt. Schließlich ist auch die Einhaltung der Regelstudienzeit am Institut mit 95% im Diplom bzw. 86% im Magisterstudiengang (Anteil der Studenten, die sich in der Regelstudienzeit befinden an allen immatrikulierten Studenten) noch relativ gut möglich. Zwar beträgt die durchschnittliche Studiendauer 12,3 Fachsemester (Diplom) bzw. 12,4 Fachsemester (Magister). Bis zur Zwischenprüfung beträgt sie aber nur 4,4 (Dipl.) bzw. 4,1 FS (MA), womit für die Studierenden gute Voraussetzungen vorliegen, das Studium in kurzer Zeit abzuschließen – wenn sie es wünschen. Leider sind dies, da aus den 2001 veröffentlichten Lehrberichten zum Studienjahr 1999/00 entnommen, nicht die aktuellsten statistischen Daten. Daten zum Studienjahr 2001/02 werden voraussichtlich erst in Zusammenhang mit dem Lehrbericht im Sommer 2003 vorliegen. Eine Ausnahme bilden lediglich die Studenten- und Studienanfängerzahlen, die zusätzlich angefordert und aktualisiert wurden, so, dass die Betreuungsrelationen aktueller sind.

Als Stärke und gute Voraussetzung für Verbesserungen wäre auch das bei der überwiegenden Anzahl der Lehrenden als sehr hoch eingeschätzte Engagement in der Lehre herauszustellen. Auch die Fähigkeiten zur Vermittlung fachlicher Zusammenhänge, die Prüfungsvorbereitung und das Studiumfeld sind im Bundesvergleich bereits Stärken, allerdings steckt hier wohl auch noch Potential. Deutlich ausbaufähig ist die Erläuterung von Prüfungsergebnissen. Aber da die Dozenten nach Einschätzung der Studierenden auch nicht seltener als an anderen Hochschulen auf Anregungen und Vorschläge eingehen, könnte vielleicht schon häufigeres freundliches Nachfragen seitens der Studierenden genügen. In einem Klima gegenseitigen Aufeinander-Eingehens, welches auch ohne zusätzliche Ressourcen herstellbar sein sollte, könnte die Soziologie der TU Dresden sehr wahrscheinlich noch um einiges bessere Einschätzungen erzielen. Dann hat es auch nur noch geringen

Einfluss, welche Kriterien letztlich im Einzelnen zur Anwendung kommen, zumal diese oft kaum beeinflussbar sind.

Dies zeigt auch eine kleine zeitlich parallel zur 2. Erhebung durchgeführte zusätzliche Befragung mittels CHE-Fragebogen (2001) mit der beim CHE (Centrum für Hochschulentwicklung) für Auswertungen festgelegten Mindestanzahl von 15 Studierenden, der ein deutlich breiteres Spektrum an Kriterien zur Situation von Lehre und Studium erfasst (vgl. Müller-Böling/Hornbostel/Giebisch/Buhr 2000). Aufgrund der nach unseren Analysen für zuverlässigere Aussagen zu geringen Befragtenzahl werden trotz weitgehender Übereinstimmung der Ergebnisse vergleichbarer Fragestellungen diese hier jedoch nicht dargestellt. Mit der 2002 erfolgten Veröffentlichung des bundesweiten CHE-Studienführers u.a. zur Soziologie stehen zuverlässigere Befragungsergebnisse zur Verfügung.

Zur detaillierteren Einschätzung der Didaktik, des Dozentenengagements, der Motivation der Studierenden und zu einigen wichtigen Rahmenbedingungen der Lehrveranstaltungen stehen dem Institut außerdem wie auch für vorhergehende Semester die Ergebnisse der regelmäßig mit großen Befragtenzahlen durchgeführten Lehrveranstaltungsbewertung zur Verfügung, wobei als Vergleichsbasis für die Lehrveranstaltungen auch Institutsergebnisse berechnet wurden. Diese wiederum lassen sich einordnen in eine hochschulübergreifende Auswertung (vgl. Krempkow/Winter 2000). Die bisher nicht mögliche und sehr wünschenswerte Analyse dieser Ergebnisse über einen längeren Zeitraum ist aber wohl erst in einigen Jahren möglich, vorausgesetzt, dass über einen längeren Zeitraum vergleichbare Befragungsinstrumente eingesetzt werden.

Ergänzend wäre künftig auch ein Langzeitvergleich von Absolventeneinschätzungen zur Situation von Lehre und Studium mit im Abstand von 3-6 Jahren durchzuführenden Absolventenbefragungen an den jeweiligen Studienrichtungen möglich (vgl. Lenz/Wagner 2001; vgl. Krempkow 2000, 2001, 2002, 2003).

## 7. Literaturverzeichnis

- Bortz, J./Döring, N. (2002): *Forschungsmethoden und Evaluation*. Berlin und Heidelberg: Springer Verlag.
- Daniel, H.-D. (1998): Studentische Beurteilung von Lehrveranstaltungen. Anlage, Durchführung und Ergebnisse eines Modellprojektes an der Universität Mannheim. In: HRK (Hg.): *Evaluation und Qualitätssicherung an den Hochschulen in Deutschland – Stand und Perspektiven*. Beiträge zur Hochschulpolitik 6/1998. S. 79-104.
- Esser, H. (1997): Zweifel an der Evaluation der Lehre. In: *WiSt* (2), Heft 1, Januar 1997, S. 45-49.
- Hage, N. (1996): Studentische Urteile über die Lehr- und Studienqualität. Analysen zur Validität anhand der Indikatoren und Skalen im Studierendenurvey. Hefte zur Bildungs- und Hochschulforschung (19). Konstanz.
- Hornbostel, Stefan/Daniel, Hans-Dieter (1996): Die Studienbedingungen in der Soziologie und die Bewertung der soziologischen Lehre durch die Studenten. In: Artus, Helmut M./Herfurth, Matthias (Hg.): *Soziologielehre in Deutschland*. Opladen: Leske + Budrich.
- Hornbostel, S. (1999): Evaluation und Ranking- Führen sie zu mehr Transparenz und Vergleichbarkeit? In: HRK (Hg.) (1999): *Viel Lärm um nichts? Evaluation von Studium und Lehre und ihre Folgen*. Tagung an der Universität Rostock vom 6. bis 8. September 1998. Beiträge zur Hochschulpolitik 4/1999. Bonn. S. 81-95.

- Hornbostel, S. (2001): Hochschulranking: Beliebigkeit oder konsistente Beurteilungen? Rankings, Expertengruppen und Indikatoren im Vergleich. In: Müller-Böling, Detlef/Hornbostel, Stefan/Berghoff, Sonja (Hg.): Hochschulranking – Aussagefähigkeit, Methoden, Probleme. Gütersloh: Verlag Bertelsmann Stiftung.
- Krempkow, René (1999): Ist gute Lehre meßbar? Untersuchungen zur Validität, Zuverlässigkeit und Vergleichbarkeit studentischer Lehrbewertungen (Edition Wissenschaft; Reihe Sozialwissenschaften, Bd. 101). Marburg: Tectum Verlag.
- Krempkow, R./Grimm, H. (1999): Die Bewertung der Studienbedingungen am Institut für Soziologie der TU Dresden aus Sicht der Studierenden. Institut für Soziologie TU Dresden, Mai 1999. In: Weiße Flecken in Sachsens Hochschullandschaft, Teil 3.
- Krempkow, R. (2000): Absolventenstudien Maschinenwesen und Mathematik/ Naturwissenschaften TU Dresden 1999 (Präsentationsmappen). Universitätsmarketing TU Dresden, April 2000.
- Krempkow, R./Heldt, M. (2000a): Was kam raus? – Eine erste Auswertung der Studentenbefragungen zur Qualität der Lehre an der TU Dresden. Hg.: Prorektor Bildung/ Dezernat Akademische Angelegenheiten.
- Krempkow, R./Heldt, M. (2000b): Das Lehrevaluationsverfahren an der TU Dresden. Entwicklung, Erfahrungen und Ergebnisse. Vortrag anlässlich des 1. Ingenieurpädagogischen Kolloquiums an der TU Dresden am 10. Februar 2000 (zu finden über „Publikationen“).
- Krempkow, R./Heldt, M. (2000c): Einflussfaktoren auf studentische Lehrbewertungen an sächsischen Hochschulen. In: Arbeitskreis hochschulpolitische Öffentlichkeit beim StuRa der Universität Leipzig (Hg.): hochschule ost Nr. 1/2 2000. Leipziger Beiträge zu Hochschule/Wissenschaft.
- Krempkow, R./Heldt, M. (2000d): Bericht über die Ergebnisse der Studentenbefragungen zur Qualität der Lehrveranstaltungen im Wintersemester 1999/2000. Studentisches Evaluationsbüro Sachsen – SES (zu finden über „Publikationen“).
- Krempkow, R./Winter, J. (2000): Bericht über die Ergebnisse der Studentenbefragungen zur Qualität der Lehrveranstaltungen im Sommersemester 2000. Gegenüberstellung der Ergebnisse Sommersemester 1999 bis Sommersemester 2000. Studentisches Evaluationsbüro Sachsen – SES.
- Krempkow, R. (2001): Dresdner Absolventenstudien 2000: Wirtschaftswissenschaften/Elektrotechnik. Abschlußbericht der Befragung von Absolventen der TU Dresden zum beruflichen Verbleib und zur retrospektiven Bewertung der Studienqualität. TU Dresden, Institut für Soziologie, Lehrstuhl für Mikrosoziologie, auch im Internet: <http://www.tu-dresden.de/phfs/lenz/fo/Absolventenstudie.html>
- Krempkow, R./König, K./Winter, J. (2001): Studienführer Sachsen: Wirtschafts-/Rechts-/Geistes- und Sozial- sowie Ingenieurwissenschaften und Informatik. Gegenüberstellung von Daten aus den Lehrberichten der Hochschulen und ausgewählten Ergebnissen studentischer Lehrbewertungen. TU Dresden, Institut für Soziologie und Institut für Politikwissenschaft im Internet: [www.sz-online.de/ranking](http://www.sz-online.de/ranking).
- Krempkow, R. (2002): Die Einschätzung der Studienbedingungen durch Studierende am Institut für Soziologie der TU Dresden 2002 und 1999. Institut für Soziologie TU Dresden.
- Krempkow, R./Dettmer, A./Eberhardt, J./Schmidt, S. (2003): Dresdner Absolventenstudien 2002: Bauingenieurwesen/Medizin. Abschlußbericht zur Befragung der Absolventen der TU Dresden zum beruflichen Verbleib und zur retrospektiven Bewertung der Studienqualität. TU Dresden, Institut für Soziologie, Lehrstuhl für Mikrosoziologie.
- Kromrey, H. (1996): Von der Lehrevaluation zur Entwicklung von Lehrqualität. Eine Bilanz bisheriger Evaluationsverfahren und ein Ausblick. Langfassung des Vortrags auf dem 28. Kongreß der Deutschen Gesellschaft für Soziologie in Dresden am 9. 10. 1996, zu beziehen über: FU Berlin, Institut für Soziologie.
- Lehrberichte TU Dresden 1996/97 bis 1999/00: (Lehrberichte müssen in Sachsen generell in den Hochschulbibliotheken für die Öffentlichkeit zur Einsicht bereitgehalten werden.)
- Lenz, K./Wagner, B. (2001): Dresdner Absolventenstudien 2000. Erziehungswissenschaftliche Fakultät, Philosophische Fakultät, Sprach- und Literaturwissenschaftliche Fakultät, Abschlußbericht. In: Institut für Soziologie TU Dresden, Lehrstuhl für Mikrosoziologie (Hg.): Ar-

- beitsberichte Dresdner Soziologie Nr. 6, im Internet: unter [www.tu-dresden.de/phfis/lenz/-fo/Absolventenstudie.html](http://www.tu-dresden.de/phfis/lenz/-fo/Absolventenstudie.html)
- Marsh, H.W. (1982): Validity of students' evaluations of college teaching – a multitrait multimethod analysis. In: Journal of educational psychology, Nr. 74(2), S. 264 – 279.
- Marsh, Herbert W./Roche, Lawrence A. (1997): Making Students Evaluations of Teaching Effectiveness Effective. In: American Psychologist, 11/97, S. 1187-1197.
- Müller-Böling, D./Hornbostel, S./Giebisch, P./Buhr, P. (2000): Der Studienführer 2000: Vorgehensweise und Indikatoren. In: CHE (Hg.): Arbeitspapier Nr. 22.
- Rindermann, H. (2001): Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen. Mit einem Beitrag zur Evaluation computerbasierten Unterrichts. Landau: Empirische Pädagogik.
- Rindermann, H./Kohler, J. (2003): Lässt sich die Lehrqualität durch Evaluation und Beratung verbessern? Überprüfung eines Evaluations-Beratungs-Modells. In: Psychologie in Erziehung und Unterricht, 50 (1), S. 71-85.
- Schwarzkopf, J. (2001): Studiengänge auf dem Prüfstand. Wie Studierende ihr Studium bewerten. (Diplomarbeit). Institut für Soziologie der Universität Potsdam.
- Spiel, C./Gössler, M.P. (2000): Zum Einfluss von Biasvariablen auf die Bewertung universitärer Lehre durch Studierende. In: Zeitschrift für Pädagogische Psychologie, 14 (1), S. 38-47.
- Teichler (2002): Potentiale und Erträge von Absolventenstudien. In: Sozialwissenschaften und Berufspraxis (SuB), 25. Jg., Heft 1/2 2002, S. 9-32.

## 8. Anhang

### 8.1 Auszug aus Krempkow/König/Winter 2001: Soziologie<sup>5</sup>

	Studenten- zahl *	Studien- anfänger *	Studien- dauer Semester **	Studier- barkeit Prozent **	Durchschnitts- noten **	Betreuung Studenten je Professor	Absolventen- quote Abschlüsse nach 5 J. Prozent
<b>UNIVERSITÄTEN (Dipl.)</b>							
TU Dresden <sup>4</sup>	499 ↻	68 ↘	12,3 ↘	95 ↘	1,9 ↻	100 ↘	17
TU Chemnitz <sup>4</sup>	297 ↻	66 ↗	10,2 ↗	94 ↘	2,7 ↗	59 ↘	3
Uni Leipzig <sup>4</sup>	526 ↗	90 ↘	10,9 ↗	89 ↗	2,0 ↘	66 ↗	32
<b>UNIVERSITÄTEN (MA)</b>							
TU Dresden <sup>4</sup>	499 ↗	68 ↘	12,4 ↗	86 ↗	2,4 ↗	100 ↘	17
TU Chemnitz <sup>4,9</sup>	297 ↗	66 ↗	10,8 ↗	95 ↗	2,3 ↗	59 ↘	18
Uni Leipzig <sup>4</sup>	526 ↗	90 ↘	9,8 ↗	93 ↗	2,5 ↗	66 ↗	5

Die Grafiken: Pfeile und pure Zahlen, doch was steckt dahinter? Es ist ein Spiegelbild der sächsischen Hochschulen. Grundlage bilden die 2000/2001 veröffentlichten Lehrberichte. Die Pfeile geben den Trend an. Sie beziehen sich auf die Werte vom Jahr zuvor. Pfeile nach oben bedeuten Verbesserungen, Pfeile nach unten in jedem Falle Verschlechterungen. Bei geringen Veränderungen (fünf Prozent) wurde kein Trend ausgewiesen. Bei der Studiendauer, den Abschlussnoten und der Betreuung bedeuten höhere Zahlen gegenüber dem Vorjahr eine Verschlechterung. Studier-

5 Unter o.g. Internetadresse [www.sz-online.de/ranking](http://www.sz-online.de/ranking) ist unter dem Link „Studienführer“ eine ausführliche wissenschaftliche Dokumentation der Datengrundlage und Methodik zu finden.

barkeit gibt den Anteil der Studenten an, die sich noch in der Regelstudienzeit befinden, im Verhältnis zur Gesamtstudentenzahl. Je geringer dieser Wert, desto schlechter. Die Absolventenquote taucht erstmals in dieser Gegenüberstellung auf. Hier wird die Zahl der Studenten verglichen, die vor fünf Jahren ihr Studium begonnen hatten, mit jener, die heute ihren Abschluss machen. Studienabbruch, Hochschulwechsel (Weggang, Zuzug), Fachwechsel und Abbruch beeinflussen diesen Wert.

## 8.2 Soziologie – Diplom und Magister an Universitäten

In Dresden und Chemnitz nimmt die Zahl der Studierenden der Soziologie gegenüber dem Vorjahr zu. In Leipzig stagniert die Gesamtzahl noch, die Zahl der Studienanfänger nimmt jedoch ab. Dresden hat von allen drei Standorten das ungünstigste Betreuungsverhältnis und die längste Studiendauer. Zwar wird das Notenspektrum bei den Abschlussprüfungen offenbar ausgeschöpft. Die eigentlichen Hürden des Studiums scheinen jedoch nicht in der Abschlussprüfung zu liegen: Diese werden in Sachsens Geistes- und Sozialwissenschaften an allen Unis von allen Teilnehmern bestanden, daher wurde dies hier nicht ausgewiesen. Allerdings zeigt die Relation zwischen Anfängern 94/95 und Absolventen 98/99, dass weit über die Hälfte aller Studienanfänger sich gar nicht erst zur Abschlussprüfung angemeldet, Hochschule oder Studienfach gewechselt, oder das Studium ganz aufgegeben haben.