

Verlässliche Evaluationsergebnisse für Qualität und Renommee der Ausbildung

Bericht 2010 der Expertengruppe Evaluation und Qualität der Fachhochschulen für den öffentlichen Dienst

Management Summary

Evaluation verbessert die Lehre, wie eine empirische Untersuchung prominenter Bildungsforscher nachgewiesen hat, vor allem dann, wenn sie mit Beratungs- und Unterstützungsangeboten kombiniert wird. Und Evaluation ist unverzichtbar für die Erfüllung des Auftrages der Hochschulen, wie der Wissenschaftsrat in seinen Empfehlungen 2008 ausgeführt hat.

Um die Leistungsfähigkeit der Evaluation zu sichern, hat die Expertengruppe

- die bisher entwickelten Instrumente der Evaluation anhand vorliegender Forschungsergebnisse und Praxiserfahrungen überprüft,
- ergänzende Empfehlungen für den Einsatz der Evaluation und insbesondere die Nutzung der Befragungsergebnisse erarbeitet

und legt diese Ergebnisse hiermit als Bericht 2010 vor.

Die Auswertung der vorliegenden Fachbeiträge bestätigt, dass die von der Expertengruppe entwickelte Studierendenbefragung

- fachlichen Standards entspricht,
- optimal für die Belange der Fachhochschulen für den öffentlichen Dienst geeignet ist,
- die einzigartige Möglichkeit des Vergleichs und damit des Lernens von anderen Hochschulen mit vergleichbaren Studiengängen bietet.

Die Expertengruppe hat das Instrumentarium zur Auswertung von Evaluationsergebnissen weiter entwickelt, weil hier ein besonderer Bedarf gesehen wird. Der mit den Befragungen entstehende Aufwand ist nur zu rechtfertigen, wenn der Informationsgehalt der Befragungsergebnisse auch ausgeschöpft wird und die für die Weiterentwicklung der Ausbildung erforderlichen Impulse liefert. Um dies zu gewährleisten, hat die Expertengruppe die fachlichen Anforderungen an

Online-Quelle dieses Berichts:

http://www.verwaltungsmanagement.info/eva/ExpGr_EvaluationBericht_2010.pdf

das Verfahren ebenso wie die statistisch-methodischen Standards beschrieben und Standards für eine aussagekräftige Aufbereitung der Ergebnisse entwickelt.

Die Expertengruppe hat weiter untersucht, wie das in den Mitgliedshochschulen verfügbare Wissen zum Thema Evaluation für alle nutzbar gemacht werden kann, insbesondere durch Nutzung „kollaborativer Programme“ des Web 2.0. Sie erleichtern Wissensmanagement, weil wesentliche Inhalte durch die Nutzerinnen und Nutzer selbst gepflegt werden können. Dennoch ist ein solches System nicht ohne verantwortliche Betreuung lebensfähig, die etwa 10% einer Vollzeitstelle (vier Zeitstunden wöchentlich) erfordert.

Die Expertengruppe macht erneut darauf aufmerksam, dass es problematisch ist, Leistungsbezahlung zu einem wesentlichen Teil nach den Ergebnissen der studentischen Lehrveranstaltungsbeurteilung zu gewähren. Damit wird die Evaluation verrechtlicht: Sie muss so gestaltet werden, dass alle Lehrkräfte die gleichen Chancen für leistungsbezogene Vergütungsbestandteile haben, gleichzeitig wird die Beachtung sozialwissenschaftlicher Standards bei der Interpretation der Daten Voraussetzung für die Rechtmäßigkeit personalrechtlicher Entscheidungen.

Mitglieder der Expertengruppe 2009/2010

Frank Hochapfel	FH Bund
Sibylle Kraheck-Brägelmann	NRW
Burkhardt Krems	FH Bund
Silke Lautenbach	Schleswig Holstein
Beate Nolte-Gehlen	Bad Münstereifel
Jaqueline Reichardt	Thüringen
Jürgen Rohdenburg	Bremen
Gabriele Schaa	Hessen
Matthias Ulbrich-Herrmann	NRW

Kontaktadresse

gabriele.schaa@vfh-hessen.de

Inhaltsverzeichnis	Seite
Management Summary	I
Verzeichnis der Abbildungen und Tabellen	IV
1 Stand und Perspektiven der Evaluation	1
1.1 Ausgangslage	1
1.2 Empfehlungen des Wissenschaftsrates	1
1.3 Empirische Ergebnisse	2
Studentische Lehrevaluation wirkt deutlich nur in Kombination mit Beratung	2
Benchmarking – Evaluationsergebnisse im Vergleich	3
1.4 Evaluation und leistungsorientierte Besoldung / Vergütung	4
1.5 Ressourcen für Evaluation und Qualitätsmanagement	5
2 Evaluation als Lernprozess	6
2.1 Evaluation folgt wissenschaftlichen Kriterien	6
2.2 Leitbild und eine „Kultur der Evaluation“	6
3 Der veranstaltungsbezogene Fragebogen	7
3.1 Zur Konstruktion des Fragebogens	7
3.2 Zur Bewährung der Studierendenbefragung	9
4 Fragen zum Verfahren	11
4.1 Was wird überhaupt gemessen?	11
4.2 Qualität der Daten	11
4.3 Verwertung der Daten	13
4.4 Befürchtung von Nachteilen	14
4.5 Evaluation des Curriculums	15
4.6 Evaluation der Evaluation	16
5 Statistisch-methodische Probleme	16
5.1 Kennzahlen	16
5.2 Interpretation von Mittelwerten	17
5.3 Streuung und ihre Bedeutung	19
5.4 Zustimmungswerte (Ja%-Werte) als verdichteter Kennwert	22
5.5 Darstellung von Ergebnissen	23

5.6	Benchmarking	26
6	Wissensmanagement	28
6.1	Ausgangslage	28
6.2	Empfehlungen	29
	Literaturverzeichnis	31
	Anhang: Empfehlungen des Wissenschaftsrates 2008 (Auszug)	33

Verzeichnis der Abbildungen und Tabellen

Abbildungen

Abbildung 1	Streuungen um den Mittelwert	21
Abbildung 2	Zusammenhang von Ja%-Werten zum Mittelwert	23
Abbildung 3	Auswertungsprofil – Beispiel aus dem Fach Polizei- und Verwaltungsrecht (PoIVR)	24

Tabellen

Tabelle 1	Mittelwerte in den Fachgebieten (ca. 35.000 Datensätze aus 7 Fachhochschulen) 2004	26
Tabelle 2	Vergleich der Ergebnisse – FB Polizei und FB Verwaltung 2003/2004 (Vergleich mit anderen Fachhochschulen) und 2009 VFH Hessen	26
Tabelle 3	Vergleich der Ergebnisse – FB Polizei 2003/2004 (Vergleich mit anderen Fachhochschulen) und Ergebnisse VFH Hessen 2004 und 2009	27
Tabelle 4	Vergleich der Ergebnisse – FB Verwaltung 2003/2004 (Vergleich mit anderen Fachhochschulen) und VFH Hessen 2009	27

1 Stand und Perspektiven der Evaluation

1.1 Ausgangslage

Die Ausbildung des gehobenen Dienstes ist eine wesentliche Einflussgröße auf die Qualität der öffentlichen Verwaltung. Pro Absolventin oder Absolvent ist sie eine Investition von mehr als 100.000 Euro.

- Deshalb hat die Evaluation einen hohen Stellenwert. Sie muss dazu beitragen, dass die Ausbildung (Auswahl der Studierenden, Studium, Praktika, Prüfung) erfolgreich qualifiziert (Qualifikationserfolg) und
- die Investitionen rechtfertigt (Investitionserfolg).

Auch wenn die Zahl der Studienbewerber bis 2025 auf einem hohen Niveau sein wird, kommt es auf die Attraktivität der Ausbildung an, um eine ausreichende Zahl von möglichst gut qualifizierten Bewerberinnen und Bewerbern zu gewinnen. Dafür ist das Renommee der Ausbildung wichtig: Die Bewerberinnen und Bewerber müssen mit der Ausbildung die Erwartung verbinden, dass sie ihr Potenzial entfalten können und ihnen eine zukunftssichere Qualifikation vermittelt wird, die ihnen lebenslanges Lernen ermöglicht. Auch dazu muss die Evaluation beitragen.

1.2 Empfehlungen des Wissenschaftsrates

Der Wissenschaftsrat hat 2008 "Empfehlungen zur Qualitätsverbesserung von Lehre und Studium" veröffentlicht.¹ Sie sind Anlass, auch die Evaluation an den Fachhochschulen für den öffentlichen Dienst (FHöD) zu überprüfen, insbesondere auf die beiden folgenden Aspekte:

- Der Wissenschaftsrat macht in seinen Empfehlungen darauf aufmerksam, dass die Einschätzung der Leistung in der Lehre nur möglich ist, wenn nicht nur das Ergebnis (Output), sondern auch die Ausgangsbedingungen (Input) berücksichtigt werden. Erst der Vergleich ergibt die Leistung der Ausbildung als Qualifikationszuwachs.
- Er betont die Bedeutung, die der eigenen Aktivität der Studierenden zukommt: „Lehrende sollten selbstorganisiertes Lernen fördern und die Studienprozesse auf die Aneignung von fachlichen sowie überfachlichen Kompetenzen ausrichten. Eigeninitiative und Eigenverantwortung der Stu-

¹ Wissenschaftsrat 2008, Online-Quelle: http://www.exzellente-lehre.de/pdf/empfehlungen_zur_qualitaetsverbesserung_von_lehre_und_studium_2008.pdf (12.02.2010).

dierenden sollten gleichermaßen gefördert und eingefordert werden. Ein solches Rollenverständnis von Lehrenden und Lernenden ist zu flankieren mit Veranstaltungsformen, die förderliche Lehr- und Lernsituationen schaffen und das aktive Lernen unterstützen.“ (S. 11)

Beide Aspekte werden durch Evaluation bisher kaum oder gar nicht abgedeckt. In der derzeitigen Form ist sie nicht geeignet, Ausmaß und Erfolg selbstorganisierten Lernens zu erfassen und insoweit eine Rückmeldung zu liefern, auch zu dem Beitrag, den die Studienbedingungen und die Lehrangebote dazu leisten. Und es fehlen die Voraussetzungen dafür, den Input zu erfassen und mit dem Output zu vergleichen, um die Leistung der Hochschule zu beurteilen.

Die Expertengruppe empfiehlt, die Evaluation entsprechend den Empfehlungen des Wissenschaftsrates zu erweitern.

1.3 Empirische Ergebnisse

Studentische Lehrevaluation wirkt deutlich nur in Kombination mit Beratung

Rindermann, ein renommierter Bildungsforscher, hat gemeinsam mit anderen empirisch untersucht, ob die Evaluation durch die studentische Lehrveranstaltungsbeurteilung einen nachweisbaren Effekt auf die Qualität der Lehre hat². Hintergrund sind Untersuchungsergebnisse aus den USA, nach denen sich die bloße Rückmeldung der Ergebnisse an die Lehrkräfte als wenig wirksam erwiesen hat.

Die empirische Überprüfung an einer deutschen Ausbildungsstätte für Sprachtherapeuten und an einer Fachhochschule³ kommt zu gleichen Ergebnissen. Sie bestätigt die Wirksamkeit einer Kombination von Rückmeldung durch Evaluation und gezielter Beratung, und zwar nicht nur auf die Bewertung durch die Studierenden, sondern zum Teil auch auf Examensergebnisse.⁴

Die Autoren fassen die Ergebnisse wie folgt zusammen:

„In Übereinstimmung mit einigen früheren Untersuchungen zeigte sich, dass Evaluation allein kaum Wirkung hat, eine deutliche Verbesserung aber erreicht wurde, wenn die Evaluation mit Beratung kombiniert wurde.“

² Rindermann/Kohler/Meisenberg 2007, Online-Quelle: <http://www.uni-graz.at/pslgcwww/rindermann/publikationen/07IntJourAD.pdf> (12.02.2010).

³ Dresel/Rindermann/Tinsner 2007, S. 193-204.

⁴ Dresel/Rindermann/Tinsner 2007, S. 76.

Und weiter heißt es dort:

„Feedback über die Wirkung der Lehre allein, ohne Beratung die es ermöglicht, die eigene Lehre und Beispiele für besseres Lehren zu reflektieren, ergibt keine messbaren Verbesserungen (...). Aber auch Beratung ohne Evaluation ist nicht wirksam, weil die Lehrkräfte selten ihre eigenen Stärken und Schwächen wahrnehmen.“⁵

Besonders profitierten die Lehrkräfte, die bisher weniger gut bewertet worden waren, auch wenn sie nicht das Niveau der schon früher gut bewerteten Lehrkräfte erreichten: Die Rangfolge der Lehrkräfte blieb erhalten.⁶

Die Autoren verweisen auch darauf, dass Evaluation ohne Unterstützung bei der Verbesserung der Lehre Frustration und Entmutigung erzeugen und deshalb sogar einen negativen Effekt haben kann. Auch aus ethischer Verantwortung sollte deshalb Beratung angeboten werden⁷.

Die Expertengruppe empfiehlt,

- vorhandene Beratungsangebote auf ihre Wirksamkeit in Kombination mit Evaluation zu überprüfen,
- wo bisher keine Beratungsangebote bestehen, diese als Ergänzung zur Evaluation aufzubauen.

Benchmarking – Evaluationsergebnisse im Vergleich

Bisher einzigartig sind die Vergleiche von Evaluationsergebnissen über Fächer, Fachbereiche und Fachhochschulen hinweg, wie sie die Expertengruppe in ihrer Pilotstudie 2004 vorgelegt hat.

Ergebnisse sind u. a.:

- Defizite in der Lehre liegen in den Aspekten „Förderung des Verständnisses für fachübergreifende Zusammenhänge“ (Mittelwert von 2,4) sowie „Förderung des selbständigen Lernens“ (Mittelwert von 2,3).

⁵ Dresel/Rindermann/Tinsner 2007, S. 83 (eigene Übersetzung). Die Divergenz zwischen Selbst- und Fremdeinschätzung ist empirisch belegt, vgl. Rindermann/Kohler/Meisenberg 2007, S. 83: „Few lecturers assess quality criteria such as structure and workload adequately. Therefore the perspective of students and external people, such as counsellors, is required.“

⁶ Dresel/Rindermann/Tinsner 2007, S. 82.

⁷ Vgl. Dresel/Rindermann/Tinsner 2007, S. 83.

- Die Bewertung der Lehrenden ist sowohl vom Interesse am jeweiligen Fach als auch von der Einschätzung der Berufsrelevanz des Faches durch die Studierenden abhängig.
- Das Interesse der Studierenden z. B. an rechtswissenschaftlichen Fächern ist höher als an wirtschafts- oder sozialwissenschaftlichen Fächern, was auch zu einer höheren Zufriedenheit in dem erstgenannten Bereich führt.
- Im Vergleich der Fachbereiche erreichen die Lehrenden im Fachbereich Polizei bessere Werte als in den anderen Fachbereichen.

Damit sind wichtige Einflussgrößen auf die Ergebnisse der Studierendenbefragungen identifiziert, aus denen aber bisher – soweit ersichtlich – kaum Konsequenzen gezogen worden sind. Sie sind auch für die Frage relevant, ob aus den Ergebnissen Schlüsse für die Beurteilung von Lehrkräften gezogen werden können.

1.4 Evaluation und leistungsorientierte Besoldung / Vergütung

Studierendenbefragungen sind ein unverzichtbarer Teil der Evaluation, aber in ihrer Aussagekraft begrenzt. Sie sind insbesondere keine verlässliche Beurteilung der Lehrqualität (siehe unten S. 11, Punkt 4.1).

Die Expertengruppe wiederholt ihre Empfehlung aus früheren Berichten, Evaluation nicht mit der Leistungsbezahlung zu verknüpfen. Die Verknüpfung verrechtlicht die Evaluation und kann erhebliche Probleme rechtlich einwandfreier personeller Entscheidungen verursachen:

- Welche fachlichen Standards der Evaluation (der Datenerhebung, Rücklaufquote, Schutz der Anonymität, statistische Auswertung, Verwendung von Mittelwert und Streuungsmaß usw.) sind rechtlich gefordert?
- Wie muss die Evaluationsplanung gestaltet sein, damit die Lehrenden angemessene Möglichkeiten haben, gute Bewertungen zu erhalten?
- Welche Zielvorgaben für Evaluationsergebnisse dürfen formuliert werden?
- Welche Vergleiche mit anderen Lehrkräften dürfen verwendet werden?

Damit werden Anlass, Inhalte, Häufigkeit, Verfahren der Durchführung bis hin zur Rücklaufquote und Interpretation der Daten möglicher Gegenstand gerichtlicher Auseinandersetzung.⁸

Bei dieser Verwendung kann Evaluation nicht mehr nach der Funktion als Teil von individuellen und organisatorischen Lernprozessen und der Unterstützung von Managemententscheidungen gestaltet werden und diese Funktion nicht mehr oder nur eingeschränkt erfüllen. Es gilt in dieser Hinsicht entsprechend, was das Bundesverfassungsgericht im Volkszählungsurteil zu Verknüpfung und Abgleich der Daten aus der Volkszählung mit den Daten der Melderegister ausgeführt hat.⁹

1.5 Ressourcen für Evaluation und Qualitätsmanagement

Evaluation, systematisch, methodisch korrekt und mit transparenten Ergebnissen betrieben, verursacht Kosten. Die Fehlleitung von Ressourcen wäre aber viel teurer. Denn die Ausbildung für den öffentlichen Dienst kostet pro Absolvent/in mehr als 100.000 Euro, stellt also eine erhebliche Investition dar. Dabei sind die Personalkosten der Studierenden für mehr als 70% der Kosten verantwortlich, weniger als 30% sind Kosten der Lehre. Alle Investitionen in die Qualität der Lehre, auch durch systematische Evaluation, verändern die Kosten der Ausbildung insgesamt also nur geringfügig, tragen aber zum Ergebnis wesentlich bei – und sind schon deshalb im Zweifel „wirtschaftlich“.

Selbst wenn die Evaluation nur dazu beiträgt, Studiendauer, Durchfall- und Abbrecherzahlen geringfügig zu verringern, macht sie sich bereits „bezahlt“, weil die eingesparten Kosten ihren Aufwand überkompensieren. Hinzu kommen die Vorteile durch bessere Qualität der Ausbildung (höhere Motivation, bessere Lernerfolge, Renommee der Ausbildung usw.).

Wirksam Sparen heißt deshalb: alle Erkenntnismöglichkeiten über eine erfolgreiche Ausbildung ausschöpfen, d. h. insbesondere: eine systematische, professionelle Evaluation. Auch die Forderungen des Wissenschaftsrates (siehe oben S. 1) werden ohne eine Professionalisierung der Evaluation nicht zu erfüllen sein.

Eine Professionalisierung bedeutet, dass für die Aufgabe mindestens eine hauptamtliche Person mit sozialwissenschaftlich-empirischer Qualifikation zur

⁸ Zu den fachlichen Anforderungen an die Erhebungsinstrumente siehe Abschnitt 3 (Anforderungen an einen Fragebogen, S. 7 ff.), an die Verwertung der Ergebnisse Abschnitte 4 und 5 (S. 11 ff.).

⁹ BVerfGE 65,1.

Verfügung steht, die auch die erforderliche Unabhängigkeit hat, um ihrer besonderen Vertrauensposition gerecht werden können. Darüber hinaus müssen die benötigten Ressourcen für die verwaltungsmäßige und technische Durchführung der Evaluation bereitgestellt werden, damit Routinetätigkeiten nicht die Kapazitäten für die konzeptionelle und inhaltliche Arbeit erschöpfen.

Die Expertengruppe empfiehlt, Organisation und Ausstattung der Evaluation daraufhin zu überprüfen, ob sie die formulierten Anforderungen erfüllen kann.

2 Evaluation als Lernprozess

Evaluation entfaltet ihre positiven Wirkungen nur im Rahmen einer angemessenen Grundhaltung der Beteiligten und als Teil organisationaler Lernprozesse.

2.1 Evaluation folgt wissenschaftlichen Kriterien

Evaluation umfasst die gezielte wissenschaftliche Untersuchung von Maßnahmen und Instrumenten der Praxis, deren Messung und die Bewertung der Auswirkungen ihres Erfolges. Wie alle Bereiche empirischer Forschung müssen auch Evaluationen wissenschaftlichen Kriterien genügen. Evaluation hat die Aufgabe, objektive und valide – also gültige – und reliable, d. h. zuverlässige Ergebnisse als Entscheidungsgrundlagen zu liefern und vermutete Ursache-Wirkungszusammenhänge (Hypothesen) zu überprüfen.

Evaluation erfüllt einen gesetzlichen Auftrag und kann eine verlässliche Grundlage für Entscheidungen nur liefern, wenn sie ergebnisoffen erfolgt und den fachlichen Standards genügt.

2.2 Leitbild und eine „Kultur der Evaluation“

Wirksame Evaluation an Hochschulen braucht ein Leitbild und eine gemeinsame Evaluationskultur. Es geht um ein gemeinsames Grundverständnis von Evaluation, welches klare Rahmenbedingungen für ein respektvolles Miteinander ermöglicht, sich auf das Ziel von Evaluation konzentriert – das Ziel der nachhaltigen Verbesserungen im Lehr- und Lernprozess – und Evaluation einbindet in die Gesamtstruktur einer lernenden Organisation. Ein solches gelebtes Grundverständnis kennzeichnet eine „Kultur der Evaluation“.

Voraussetzung für den Erfolg organisationaler Lernprozesse ist eine positive Fehlerkultur in der Hochschule. Fehler sind Lernchancen; es geht nicht darum,

Schuldige zu finden. Werden Schwierigkeiten wie das Nichterreichen von Zielen – so bezeichnete „Fehler“, Mängel und Misserfolg – „bestraft“, dann erzeugen solche Maßnahmen Widerstand bei den Betroffenen und/oder inneren Rückzug. Evaluation und Qualitätsmanagement als internalisierter, kontinuierlicher Lernmechanismus in der Hochschule ist dann zum Scheitern verurteilt.

Eine positive Fehlerkultur in einer Organisation setzt ein positives Menschenbild voraus, das wachsen muss und dessen erste Annahme es ist, dass die Akteure in der Hochschule allesamt gewillt sind, ihre eigene Leistungserbringung immer wieder auf den Prüfstand zu stellen und dazuzulernen. Negative Stereotypen und Vorurteile verhindern die notwendige Vertrauenskultur, ohne die es nicht gelingt, Personen für Evaluation und Qualitätsmanagement zu gewinnen, die das Vertrauen aller Beteiligten genießen.

Ein Evaluationskonzept muss deshalb auf der einen Seite größtmögliche Transparenz erlauben (Ehrlichkeit), auf der anderen Seite aber Schutzmechanismen etablieren (Einhaltung von Datenschutzregeln), die eine Vertrauenskultur ermöglichen. Die Umsetzung wissenschaftlicher Standards – Richtigkeit, Wahrhaftigkeit – gehört genauso zu einem glaubwürdigen Konzept wie eine Verankerung solcher Systeme auf breiter Basis, so dass alle Akteure sich aktiv beteiligen können (Nachvollziehbarkeit).¹⁰ Insgesamt ist es notwendig, sich bewusst zu machen, dass Evaluation viele Möglichkeiten bietet zu zählen, zu messen, zu analysieren und zu interpretieren, aber auch Risiken der Fehlinterpretation und damit der Fehlsteuerung birgt.

3 Der veranstaltungsbezogene Fragebogen

3.1 Zur Konstruktion des Fragebogens

Die Konstruktion des Fragebogens ist entscheidend für die Qualität dieses Messinstruments. Bei der Konstruktion muss man zurückgreifen auf ein Verständnis von „guter Lehre“¹¹ und Hypothesen.

Ein Fragebogen zur studentischen Lehrevaluation muss diese relevanten Aspekte und Einflussfaktoren enthalten, die untersucht werden sollen und deshalb standardmäßig erfasst werden müssen. Der Fragebogen zur studentischen Lehrevaluation, den die Expertengruppe Evaluation und Qualität empfiehlt, ist

¹⁰ Vgl. Evaluationsstandards der DeGEval – Gesellschaft für Evaluation e.V.: Nützlichkeit, Durchführbarkeit, Fairness, Genauigkeit.

¹¹ Es gibt einen gesicherten Kenntnisstand darüber, was „gute Lehre“ ist – vgl. Hilpert/Meier siehe bei olev.de.

empirisch in verschiedene Richtungen durch Itemanalysen auf seine Tauglichkeit, Qualitätsaspekte der Lehre und Zufriedenheit zu messen, geprüft.

Mittlerweile gibt es eine Vielzahl an Verfahren für die Evaluation und Qualitätssicherung von Lehrveranstaltungen. Entsprechend der Vielfalt an Konstruktionsmöglichkeiten liegen den Verfahren auch unterschiedliche Dimensionen zugrunde, die die „Qualität von Lehre“ erfassen sollen. Der 2008 empfohlene Fragebogen für die Lehrevaluation von Modulen bzw. Teilmodulen oder Lerneinheiten deckt die zentralen Bewertungsdimensionen von Lehre ab¹² und eignet sich für den horizontalen und vertikalen Vergleich sowie den Vergleich von Fachhochschulen (siehe Abschnitt 5.6).

Bei der Konstruktion des Fragebogens sind folgende „10 Regeln“ zu beachten:

1. Die Aussagen sind kurz und verständlich (einfache, klare, direkte Sprache).
2. Die verwandten Begriffe sind eindeutig, werden von allen Befragten in gleicher Weise verstanden.
3. „Unklare“ Begriffe sind definiert (z. B. „Medien“ oder „selbstständiges Lernen“).
4. Der Kontext einer Frage wirkt sich nicht auf deren Beantwortung aus.
5. Jede Aussage bezieht sich auf einen einzigen vollständigen Gedanken. Es gibt keine doppelten Stimuli (zwei Gedanken in einer Aussage).
6. Alle Aussagen sind positiv formuliert (keine doppelten Verneinungen).
7. Aussagen zielen auf Beurteilungen, die alle Befragten vornehmen können.
8. Die Antwortkategorien (5 stufige Skala) sind benannt und überschneiden sich nicht.
9. Am Ende des Fragebogens wird nach Anregungen, Vorschlägen und Kritik offen gefragt.
10. Der Fragebogen ist mit geringem Zeitaufwand auszufüllen.

¹² Vgl. Bericht der Expertengruppe Evaluation und Qualität 2008, S. 2 f.

3.2 Zur Bewährung der Studierendenbefragung

Die Expertengruppe hat verschiedentlich geäußerte Kritik am Fragebogen aufgegriffen und die aktuelle Literatur zum Instrument „Studierendenbefragung“ ausgewertet.

Fragebogen

In Hochschulen werden zahlreiche Fragebogen eingesetzt, mehrere sind auch intensiven Überprüfungen unterzogen worden und gelten insoweit als anerkannte Befragungsinstrumente, zum Beispiel das „Heidelberger Inventar“. Beim Vergleich mit diesen Fragebögen zeigt sich, exemplarisch auch beim „Heidelberger Inventar“ (HILVE),

- dass sie wesentlich umfangreicher sind (HILVE: mehr als 40 Fragen) und demnach sowohl von den Befragten als auch bei der Auswertung und Interpretation einen höheren Aufwand verursachen, dem kein entsprechender Zusatznutzen gegenübersteht.
- dass sie an der klassische Vorlesung als Großveranstaltung orientiert sind, also gerade nicht an der Veranstaltungsform, die an den Fachhochschulen für den öffentlichen Dienst überwiegt. HILVE geht davon aus, dass Lernen in der Vorlesung, durch Vortrag und Diskussion, stattfindet. Unterstützende Materialien (Skripte, Folien usw.) spielen keine Rolle, obwohl sie eine wesentliche Grundlage für das eigenständige Lernen der Studierenden sind.

Eine praktikable Evaluation muss den Aufwand in Grenzen halten, sonst ermüdet sie, wird nicht ernst genommen oder nicht mehr konsequent durchgeführt. Im Übrigen müssen auch Aufwand und Erkenntnisgewinn in einem angemessenen Verhältnis stehen.

Deshalb gibt es verschiedene Ansätze, eine solche, auf einen knappen Fragebogen reduzierte Evaluation durchzuführen.¹³ Zumbach u. a.¹⁴ untersuchen bewusst, wie die Evaluation praktikabel gestaltet werden kann, weil übliche Befragungsinstrumente zu aufwändig sind.

¹³ Universität St. Gallen hat eine „Stelle für Qualitätsentwicklung“ eingerichtet. In dem Gesamtsystem der Evaluation ist auch die Befragung der Studierenden in einem Bachelor-Studiengang vorbereitet: mit einem Fragebogen mit 6 Fragen:
<http://www.unisg.ch/org/hsg/qs.nsf/wwwPubInhalteGer/Musterfrageboegen> (10.03.2010).

¹⁴ Zumbach/Spinath/Schahn/Friedrich/Kögel 2007, S. 317-325.

Der Diskussionsstand bestätigt also die von der Expertengruppe entwickelte Lösung

- mit einem auf das Wesentliche beschränkten Fragebogen,
- der spezifischen Orientierung an den für den Qualifikationserfolg wesentlichen Faktoren,
- angepasst an die Situation der Fachhochschulen für den öffentlichen Dienst.

Sie ist fachlich fundiert und praxisgerecht und kann deshalb weiterhin empfohlen werden.

Die Verwendung des Fragebogens der Expertengruppe eröffnet zudem die Möglichkeit des Vergleichs zwischen unterschiedlichen Studiengängen und Hochschulen, woraus sich weitere wichtige Informationen ergeben:

- für eine realistische Interpretation der Ergebnisse, zum Beispiel differenziert nach Fächern und Fachbereichen,
- für Hinweise auf mögliche Problemaspekte,
- für „gute Lösungen“, von denen Impulse für die Weiterentwicklung der eigenen Studiengänge abgeleitet werden können.

Antwortskala

Ein fachlicher Kritikpunkt ist wiederholt die bei der Befragung verwendete fünfstufige Antwortskala: Es wird gefordert, nur eine gerade Zahl von Antwortmöglichkeiten zuzulassen, um eine jeweils positive oder negative Antwort zu erzwingen und die „Tendenz zur Mitte“ zu vermeiden.

Diese Kritik verkennt die Funktion der Evaluation. In der Marktforschung mag es ein legitimes Anliegen sein, eine positive oder negative Bewertung zu erzwingen. Hier geht es um die Abbildung vorhandener Einschätzungen, so dass die Skala sich daran zu orientieren hat, ob diese Einschätzungen ausreichend genau abgebildet werden. Wenn Befragte die Einschätzung „teils/teils“ – also weder eine negative noch eine positive Einschätzung zu der jeweiligen Frage – haben, muss es ihnen möglich sein, so zu antworten.

Die fünfstufige Antwortskala ist also für die Evaluation fachlich richtig.

4 Fragen zum Verfahren

4.1 Was wird überhaupt gemessen?

Die Lehrevaluation misst subjektive Einschätzungen von Studierenden und damit weder die objektive Qualität der Lehrveranstaltungen noch den Lernerfolg. Das belegen auch Forschungsergebnisse.¹⁵ Die statistische Auswertung der subjektiven Einschätzungen ermöglicht Rückschlüsse auf Aspekte, die für die Lehrqualität relevant sein können. Diese müssen aber aus den Daten erschlossen und durch weitere Analysen gefestigt werden.

„Das sind ja nur Meinungen, alles subjektiv“ – so reagieren gelegentlich Lehrende auf Ergebnisse aus Studierendenbefragungen. Die Subjektivität wird als Makel der Ergebnisse betrachtet. Ein Argument dagegen: Es stimmt, es handelt sich um subjektive Einschätzungen, die allerdings verhaltensrelevant werden können. Auch wenn ein Ergebnis vielleicht „objektiv falsch“ ist: Wenn die Studierenden die Vorbereitung des Lehrenden nicht erkennen, ist es irrelevant, ob objektiv eine Vorbereitung stattgefunden hat. Wenn eine Lehrkraft den Eindruck hinterlässt, nicht vorbereitet zu sein, hat dies Einfluss auf die Lernmotivation und auf die Aktivierung der Studierenden. Und wenn sie aktiv werden, werden sie eher lernen und Lernerfolg haben.

4.2 Qualität der Daten

Ein Kriterium bezüglich der Qualität der erhobenen Daten ist die Rücklaufquote. Die Rücklaufquote stellt die Anzahl verwertbarer Fragebogen in Relation zu der Anzahl versandter oder ausgegebener Fragebogen dar.

Eine hohe Rücklaufquote ist in jedem Fall anzustreben. Meist ist die Teilnahme an der Evaluation freiwillig. Um eine hohe Rücklaufquote zu erzielen, sollten die Studierenden über die Vorgehensweise und Bedeutung der Evaluation gut informiert werden. Wenn Studierende mit bestimmten Aspekten der Lehrveranstaltung unzufrieden sind, ist die Rücklaufquote erfahrungsgemäß hoch. So wie sich negative Erfahrungen schneller als positive herumsprechen, so besteht

¹⁵ Kromrey 2003 (S. 234, Online-Quelle S. 2, Fußn. 2) referiert empirische Ergebnisse bereits aus den Jahren 1994 und 1995, die folgende Hypothesen widerlegen:

- „Sie (studentische Lehrbewertungen) 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.“

Die Annahme, die studentische Lehrbewertung liefere verlässliche Aussagen über die Qualität der Lehre, sollte also als empirisch widerlegt gelten.

auch bei Studierendenbefragungen bei negativen Empfindungen ein größeres Kommunikationsbedürfnis. Wenn sie hingegen überwiegend zufrieden sind, so ist das Interesse, sich an der Befragung zu beteiligen, geringer – besonders dann, wenn eine Vielzahl von Befragungen durchgeführt wird. Sehen die Studierenden später keine Konsequenzen aus Evaluationsergebnissen, steigt die Anzahl von Verweigerern. Viele Ausfälle führen also immer zu zahlreichen Fragen und Problemen.

Werden Teilerhebungen durchgeführt – also nicht alle Studierenden befragt – fällt im Zusammenhang mit der Rücklaufquote der Begriff „Repräsentativität“. Als repräsentativ gilt eine Teilerhebung, wenn aus dem Ergebnis möglichst genau auf die Verhältnisse der Grundgesamtheit geschlossen werden kann. Die Frage, inwieweit die Ergebnisse denn repräsentativ sind, ist oft die prinzipielle Frage danach, ob ein Forschungsergebnis denn verlässlich, glaubwürdig und verallgemeinerungsfähig ist.

Das Konzept der Repräsentativität ist durchaus kritisch zu betrachten. Die „mangelnde Repräsentativität“ wird oft als Argument für Totalerhebungen angeführt, bei denen „Fehlerfreiheit“ der Ergebnisse quasi als selbstverständlich gilt. Dagegen steht bei großen Grundgesamtheiten der hohe Aufwand und die damit verbundenen Kosten; im Falle von Studierendenbefragungen ist die unweigerlich eintretende Evaluationsmüdigkeit ebenfalls ein gewichtiges Argument gegen (häufige, z. B. halbjährliche) Totalerhebungen. Denn die Evaluationsmüdigkeit führt sicher zu „Fehlern“ bei Totalerhebungen. Zudem ist die Frage der Repräsentativität verbunden mit einer „echten“ Zufallsauswahl, die oft weder sinnvoll noch machbar ist.¹⁶

Für eine gute Qualität der Daten sollten deshalb Teilerhebungen mit hoher Rücklaufquote angestrebt werden. Rücklaufquoten von 90% sind nach den Erfahrungen erreichbar, mindestens sollten sich mehr als die Hälfte der Studierenden, die zu einer Lehrveranstaltung befragt werden, beteiligen.

Um der Evaluationsmüdigkeit vorzubeugen, wird ein Evaluationsplan empfohlen. Zum Beispiel könnten in jedem Semester ein oder zwei Module evaluiert

¹⁶ „Im Falle einer ‚echten‘ Zufallsauswahl ist Repräsentativität schon von sich aus kein Gütekriterium, weil das zufällige Ziehen einer repräsentativen Auswahl in dem Sinne, dass Strukturgleichheit zwischen der Grund- und der Teilgesamtheit besteht, als ziemlich unwahrscheinlich anzusehen ist und daher fast alle Zufallsstichproben als ‚nicht repräsentativ‘ zu bezeichnen wären. Hier sollte zur Beurteilung der ‚Qualität‘ einer Schätzung auf das Konzept des Stichprobenfehlers zurückgegriffen werden bzw. sollten Formen der Stichprobenziehung verwendet werden, die eine Verkleinerung dieses Fehlers bewirken (z.B. geschichtete Stichproben)“. Von der Lippe/Kladroba 2002, Online-Dokument, S. 10.

werden.¹⁷ Die Lehrenden sollten aber innerhalb von etwa drei Jahren einen Überblick über die Zufriedenheit der Studierenden mit allen angebotenen Modulen erhalten.

4.3 Verwertung der Daten

Die Befragungsergebnisse der einzelnen Lehrveranstaltungen – erhalten

- die jeweilige Lehrkraft (Dozentin oder Dozent) und
- die oder der Qualitätsbeauftragte.

Die Ergebnisse in aggregierter (zusammengefasster) Form, die die Ergebnisse einzelner Lehrenden anonymisieren, erhalten

- die Modulverantwortlichen,
- die Abteilungsleitungen und
- die Hochschulleitung.

Wie ist jeweils mit den so bekannt gegebenen Daten zu verfahren?

Für die oder den **Lehrenden** stellt das Ergebnis zunächst einmal eine Rückmeldung dar, aus der Rückschlüsse auf zukünftiges Lehr- und evtl. Sozialverhalten gezogen werden sollten. Die Ergebnisse stellen darüber hinaus als Kommunikationsinstrument die Grundlage für die notwendige Besprechung mit den Studierenden dar.

Für **Qualitätsbeauftragte** sind die Daten dann von Bedeutung, wenn nach den Bestimmungen der Evaluationsordnung qualitätsfördernde Maßnahmen zu ergreifen sind. Wenn sich ein erheblicher Teil der Studierenden in einem oder mehreren Punkten in Bezug auf die lehrende Person unzufrieden äußert, ist es in jedem Fall notwendig, über die fachliche oder didaktische Qualität der Lehrveranstaltung und den Qualifizierungsbedarf der Lehrenden nachzudenken.

Modulverantwortliche, Abteilungs- und Hochschulleitung müssen sich mit den Ergebnissen auseinandersetzen und angemessen reagieren. Dabei ist zu berücksichtigen, dass die Ergebnisse von den Einstellungen der Studierenden geprägt sind, die aus unterschiedlichsten Gründen – z. B. „ist dieses Fach für mich praxisrelevant?“, zeitliche Lage der Lehrveranstaltung, Lehrgruppengröße,

¹⁷ Beispiel für einen Evaluationsplan im Modulbuch des Fachbereichs Verwaltung der Verwaltungsfachhochschule in Wiesbaden, Teil 1 – Das Konzept, S. 62 f. Online-Quelle: www.vfh-hessen.de.

Spannungen zwischen den Studierenden – entwickelt wurden und damit Einflussgrößen sind, auf die die Lehrenden keine oder nur geringe Einwirkungsmöglichkeiten haben.¹⁸ Dennoch wird es bei erheblichen negativen Ergebnissen notwendig sein, über das Curriculum, die Modulstruktur, die Aufteilung in Präsenz- und Selbststudium, die Rahmenbedingungen und weitere Elemente des Studiums nachzudenken.

4.4 Befürchtung von Nachteilen

Eine Befragung führt nur dann zu unverfälschten Ergebnissen, wenn die Befragten wahrheitsgemäß und offen antworten. Deshalb ist die Sicherstellung der Anonymität der Befragten unverzichtbar.

Studierende befürchten vereinzelt, dass sie Nachteile dadurch erfahren könnten, dass bei der Befragung mittels ausgehändigtem Fragebogen die Lehrenden aus der Schrift Rückschlüsse auf den Ausfüllenden ziehen könnten. Diese Befürchtung kann durch Online-Befragungen ausgeräumt werden. Sofern die Lehrenden die Originalfragebogen zurück erhalten, besteht die Möglichkeit, dass die Anonymität aufgehoben werden kann. Bei Papier- und Bleistift-Befragungen mit automatisierter Erfassung und Auswertung (z. B. EvaSys) werden die quantitativen Daten von den Anmerkungen auf der Rückseite getrennt. Allerdings werden die Beurteilungen auf der Rückseite eingescannt und eine Schrifterkennung ist so auch möglich. Da die Studierenden auf diese Möglichkeiten hingewiesen werden, können sie entsprechende Konsequenzen ziehen (z. B. Druckschrift verwenden). Die Expertengruppe empfiehlt daher generell Online-Befragungen, die keine Rückschlüsse dieser Art ermöglichen.

Lehrende könnten versuchen, Einfluss auf die Ergebnisse zu nehmen, wenn damit persönliche Konsequenzen verbunden sind: Fortsetzung des Lehrauftrages, Leistungszulage bei W-besoldeten Lehrkräften, Beförderungen, im Einzelfall sogar disziplinarische Konsequenzen. Die Expertengruppe betont deshalb erneut, dass die Verwendung der Evaluationsergebnisse zu personellen Entscheidungen dem Zweck der Evaluation widerspricht und ihre Funktion als Rückmeldung und Beitrag zur Weiterentwicklung durch Lernprozesse beeinträchtigt (siehe S. 4).

¹⁸ Das ist mehrfach empirisch untersucht worden, vgl. z. B. Fredersdorf/Lehner 2004 (S. 16), und die Korrelation z. B. mit der Beliebtheit von Fächern ist ein gesichertes Ergebnis der von der Expertengruppe selbst ausgewerteten Erhebungen (siehe S. 3). Die Antworten spiegeln auch die impliziten Theorien der Studierenden über ihr Lernen wider, geprägt durch bisherige Lernerfahrungen und Vorstellungen über und Wünsche an den Lernprozess und an Lernergebnisse

Erfahrungen zeigen, dass die Regelung, Einzelergebnisse lediglich dem Betroffenen und dem Qualitätsbeauftragten bekannt zu machen, solcherlei Befürchtungen am wenigsten aufkommen lässt.

Es ist Aufgabe des Qualitätsbeauftragten, einerseits die Studierenden darüber zu informieren, dass sachliche Hinweise zur Qualität der jeweiligen Lehrveranstaltung erbeten werden und persönliche Angriffe unterlassen werden sollen, andererseits sollten die Lehrenden über die Bedeutung der Evaluation für die Qualität der Lehre, die Chancen, die sich aus dem Feed-back ergeben informiert werden. Soweit dem Qualitätsbeauftragten bei der Auswertung der Fragebogen unsachliche Äußerungen bekannt werden, sollte den Lehrenden ein spezielles Beratungsangebot gemacht werden.

Vielfältige Erfahrungen zeigen, dass es sich dabei um Einzelfälle handelt und die große Mehrheit der Studierenden den Wert der Evaluation als Element der Qualitätssicherung erkennt und sich verantwortungsvoll an der Befragung beteiligt. Nicht unterschätzt werden sollte auch die Bedeutung des sich anschließenden Gesprächs zwischen Lehrenden und Studierenden über das Befragungsergebnis.

4.5 Evaluation des Curriculums

Mit der Einführung von Bachelorstudiengängen ist eine regelmäßige Überprüfung des Curriculums erforderlich. Dazu sind Befragungen der Lehrenden und der Studierenden, der Absolventinnen und Absolventen sowie der Abnehmer geeignet. Meist handelt es sich um quantitative empirische Forschung. Die Hochschule gewinnt einen Überblick, wie die Qualität der Lehre von ihren Kunden eingeschätzt wird.

Sollen einzelne Aspekte nachhaltig verbessert werden, sind qualitativ angelegte Instrumente besonders wirkungsvoll. Dazu gehören sogenannte Qualitätszirkel, in denen unter Moderation – z. B. des Qualitätsbeauftragten – Lehrende, Studierende, Curriculaverantwortliche und möglichst auch die Verwaltung der Hochschule – u. U. unter Vorlage der Ergebnisse schriftlicher Befragungen – über einzelne Qualitätsaspekte und die Studierbarkeit des Curriculums beraten und diskutiert wird. Die Ergebnisse werden dokumentiert und in den zuständigen Gremien mit dem Ziel der Umsetzung der Vorschläge beraten. In vertretbaren Abständen ist dann erneut über die geänderten Curricula, vor allem im Hinblick darauf, dass realistische (Lern-)Ziele festgelegt werden, zu beraten.

4.6 Evaluation der Evaluation

Das Verfahren der Evaluation ist schriftlich in einer Rechtsnorm zu regeln. Dabei ist zu berücksichtigen, dass Evaluation nicht statisch sein kann: Auch das Evaluationsverfahren muss regelmäßig evaluiert werden, insbesondere im Hinblick auf folgende Fragen:

- Liefern die Evaluationsergebnisse die gewünschten Informationen und/oder kann die Aussagekraft verbessert werden?
- Werden die Ergebnisse sinnvoll verwertet, auch für strategische Entscheidungen?
- Wird den Belangen der Betroffenen (einschließlich des Datenschutzes) Rechnung getragen?,
- Erfordern neue Erkenntnisse möglicherweise Anpassungen?
- Reichen die Kapazitäten für die Evaluation aus, damit sie ihre Funktion erfüllen kann?
- Stehen Evaluationsaufwand und Erkenntnisgewinn in einem angemessenen Verhältnis?

5 Statistisch-methodische Probleme

In diesem Abschnitt geht es um die Frage, was aus den Ergebnissen mit welcher Verlässlichkeit geschlossen werden kann und welche Fehlerquellen bei der Interpretation vermieden werden müssen. Die Expertengruppe entspricht in diesem Abschnitt dem Wunsch von Lehrenden aus den Fachhochschulen – die zum großen Teil Statistik-Laien sind – nach begrifflichen Klärungen.

5.1 Kennzahlen

Ergebnisse aus Studierendenbefragungen werden in Kennzahlen zusammengefasst. Unter einer Kennzahl soll hier eine einzige Zahl, in der eine Gruppe verschiedener aber ähnlicher Merkmalswerte zusammengefasst ist, verstanden werden. Diese Zusammenfassung wird in der Statistik als Aggregation bezeichnet. Kennzahlen bilden Eigenschaften bzw. Zustände ab. Sie werden im Rahmen der Evaluation durch die Berechnung von Mittelwerten (siehe S. 17 f.), der Streuung (siehe S. 19 f.) und durch Zustimmungswerte (Ja-%-Werte, siehe S. 22 f.) entwickelt.

Im Vergleich mit den Ergebnissen aus anderen Kursen, im Vergleich des Vorjahrs usw. regen die Kennzahlen an, über die Realität und die dokumentierten

Aspekte zu diskutieren und sich mit der Entwicklung kritisch auseinanderzusetzen. Kennzahlen ersetzen intuitive Urteile und oft pauschale, undifferenzierte Urteile – „wir sind doch ganz gut“ – durch nachprüfbare Daten. Sie sensibilisieren für Aspekte, die sonst oft nicht wahrgenommen würden. Damit ermöglichen sie die Diskussion und versachlichen sie.

Eine Kennzahl kann als Zielgröße (Soll-Wert für operative Ziele) festgelegt werden. Durch den Vergleich zwischen Soll- und Ist-Wert wird dann die Zielerreichung gemessen. Damit haben Kennzahlen auch eine Anreizfunktion, sich ständig für Verbesserungen einzusetzen. Sie erlauben präzise und herausfordernde Zielsetzungen und die Überprüfung, ob diese Ziele auch erreicht worden sind. Damit kann Fehlentwicklungen frühzeitig entgegen gewirkt werden.

Kennzahlen schaffen also eine Vergleichsbasis und erlauben damit objektive Vergleiche:

- über die Zeit:
gibt es auffällige Veränderungen zu den Vorjahren? Sind wir besser geworden? Wie ist der Trend?
- mit Soll-Werten
Haben wir unsere Ziele erreicht? Wie ist die voraussichtliche Entwicklung?
- mit anderen (Benchmarking)
Wo stehen wir im Vergleich zu anderen? Wie entwickeln wir uns im Vergleich mit anderen?

Objektive und nachprüfbare Vergleiche auch mit anderen Einrichtungen erfordern Kennzahlen, die nach dem gleichen Verfahren ermittelt wurden. Dieses System des Benchmarkings ist durch die Arbeiten der Expertengruppe realisiert worden.

5.2 Interpretation von Mittelwerten

Mittelwerte kennzeichnen die zentrale Tendenz einer Verteilung. Sie gehören zu den „Lagemaßen“ in der Statistik, die die Charakteristik einer Verteilung mit möglichst wenig Informationen beschreiben: Durch einen Wert wird charakterisiert, wo das „Zentrum“ oder der „Schwerpunkt“ einer Verteilung liegt.

Dabei gibt es verschiedene Mittelwerte, deren Berechnung auf jeweils unterschiedlichen Konstruktionsprinzipien basiert. Deshalb liegt bei den verschiedenen Mittelwerten jeweils ein anderes Verständnis davon zugrunde, was unter der „Mitte“ verstanden werden soll. Die für die Analyse der Daten aus Studie-

rendenbefragungen relevanten Mittelwerte sind der Modalwert und das arithmetische Mittel.

Der **Modalwert** – der häufigste Wert – zeigt, in welchem Bereich die meisten Einschätzungen liegen. Er beschreibt eine Verteilung gut, wenn sich eine Ausprägung deutlich hervorhebt. Der Modalwert wird nicht durch „extreme“ Werte beeinflusst – wenn z. B. wenige Studierende die Einschätzung „trifft gar nicht zu“ oder „trifft voll zu“ ankreuzen. Es kann mehrere Modalwerte geben.

Das **arithmetische Mittel** (kurz Mittelwert genannt) beschreibt die durchschnittliche Bewertung. Der Mittelwert wird berechnet, indem alle Ausprägungen addiert und die Summe durch die Anzahl der Merkmalsträger geteilt wird.¹⁹ Er kann also nur berechnet werden, wenn die Antwortmöglichkeiten codiert sind. Bei dem von der Expertengruppe vorgeschlagenen Fragebogen wird in einigen Fachhochschulen eine Codierung von 1 „trifft voll zu“ bis 5 „trifft gar nicht zu“ durchgeführt, in anderen von 0 „trifft gar nicht zu“ bis 4 „trifft voll zu“. Im ersten Fall erinnern die Einschätzungen an Schulnoten, im zweiten Fall sollen die Einschätzungen explizit nicht mit Schulnoten in Verbindung gebracht werden.

Der Mittelwert kann nur dann sinnvoll und richtig interpretiert werden, wenn man berücksichtigt, wie der Wert berechnet wird und welche Effekte auftreten können.²⁰

Ist ein Mittelwert von 2,9 ein gutes Ergebnis? Diese Frage kann nur beantwortet werden, wenn die Polung der Skala (von 0 bis 4 oder von 5 bis 1) und die Codierung klar ist. Zum Beispiel könnte ein Maßstab bei einer Kodierung von 0 = „trifft gar nicht zu“ bis 4 = „trifft voll zu“ lauten: *Ein Mittelwert über 3,0 ist ein überdurchschnittlich gutes Ergebnis, ein Mittelwert zwischen 2,5 und 3,0 ist ein gutes Ergebnis* usw. Dieser Maßstab kann zur Interpretation einzelner Evaluationsergebnissen einer Lehrkraft angewandt werden (siehe Beispiel Abbildung 3, S. 24).

¹⁹ Das arithmetische Mittel erfordert eine Intervallskala, das heißt die Messung von Variablen, bei denen die Abstände zwischen den Skalenpunkten gleich sind oder zumindest als gleich angesehen werden können. Bei der in den Fragebogen verwendeten verbalisierten 5er-Skala ist nicht von einer Gleichheit der Abstände auszugehen – es handelt sich ja nicht um ein mathematisch genaues Instrument. Das ist jedoch auch nicht notwendig für den angestrebten Zweck und so entscheiden wir uns für eine gewisse Ungenauigkeit, die jedoch eine deutliche Tendenz ausdrückt. Wer die Berechnung des arithmetischen Mittels bei dieser Skala als nicht zulässige Rechenoperation kritisiert, kann ausschließlich den Ja%-Wert (siehe S. 22 ff.), der lediglich das ordinale Skalenniveau annimmt, verwenden. In beiden Fällen sind die Ergebnisse gleichartig und ausreichend genau.

²⁰ So wird man z. B. bei der Kenntnis der mittleren Lebenserwartung im Mittelalter immer fragen müssen, wie hoch die Kindersterblichkeit gewesen ist; die – weil sie in der Tat sehr hoch gewesen ist – die Durchschnittslebenserwartung natürlich erheblich reduziert hat.

In unseren Evaluationen benutzen wir zusätzlich einen Mittelwert, der als Kennzahl für die eingeschätzte pädagogische Kompetenz einer Lehrkraft angesehen werden kann: der Mittelwert über alle Aussagen, die die eigentliche Lehre betreffen. Wenn der Kurs einer Lehrkraft (siehe S. 24) mit einem Mittelwert von 2,6 über die Aussagen 4 bis 16 gekennzeichnet wird, so kann dies insgesamt – siehe der Maßstab – als gutes Ergebnis über alle dozentenbezogenen Aussagen des Fragebogens gewertet werden.

Der Mittelwert berücksichtigt „extreme“ Werte. Aufgrund seiner Berechnung nivelliert er jedoch große Häufigkeiten an den Extremen (z. B. Werte 0 und 4) zu einem mittleren Wert zwischen 2 und 3. Genauso führen aber auch große Häufigkeiten in den Kategorien 1 oder 2 oder gleich hohe Häufigkeiten in allen 5 Kategorien zu einem Mittelwert zwischen 2 und 3. Auch die Aussagekraft des Mittelwertes ist beschränkt, wenn die Verteilung keinen eindeutigen Schwerpunkt hat.

Durch den Mittelwert allein lässt sich eine Datenmenge also nicht ausreichend charakterisieren. Benötigt wird die Information, wie die einzelnen Werte sich um den Mittelwert verteilen. Deshalb wird die Streuung zur Interpretation einer Verteilung herangezogen.

5.3 Streuung und ihre Bedeutung

Die Streuung zeigt, wie nahe die Daten insgesamt am „Zentrum“ liegen (oder auch nicht). Die Streuung stellt die Verteilung von einzelnen Werten um den Mittelwert dar. Sie misst die Unterschiedlichkeit der Einschätzungen, wie stark sich die einzelnen Einschätzungen voneinander unterscheiden und liefert damit eine wichtige Zusatzinformation zum Mittelwert. Mit der Verteilung der Häufigkeiten auf die fünf Kategorien ist zu erkennen, ob die befragte Gruppe eher homogen (die meisten Werte in ein oder zwei Kategorien) oder eher heterogen (Werte verteilen sich auf alle Kategorien relativ gleichmäßig) geantwortet hat. Ein homogenes Antwortverhalten einer Gruppe bedeutet, dass die meisten Aussagen gleich bzw. ähnlich eingeschätzt wurden. Ein heterogenes Antwortverhalten deutet dagegen darauf hin, dass die Meinungen weit auseinander gehen. Eine hohe Streuung – wenn eine Lehrveranstaltung z. B. von einem Teil der Studierenden positiv, von einem anderen Teil aber negativ bewertet wird – sollte z. B. Anlass sein genauer nach dem Grund zu fragen, um Verbesserungen für die unzufriedenen Studierenden zu erreichen, ohne die Vorteile für die begeisterten Studierenden zu gefährden.

Die Streuung zeigt auch an, wie stabil die Bewertung voraussichtlich ist: bei einer hohen Streuung – d. h. stark unterschiedlichen Antworten – ist die Zusammensetzung der Gruppe der Befragten von hohem Einfluss auf den Mittelwert und sind Veränderungen im Zeitvergleich möglicherweise nur die Folge einer anderen Zusammensetzung der Gruppe der Befragten.

Diese Unterschiedlichkeit wird in einer statistischen Kennzahl quantifiziert. Die Standardabweichung ist die gebräuchlichste Kennzahl für die Streuung.²¹

Die Berechnung der Standardabweichung – mit dem Buchstaben s abgekürzt – ist etwas kompliziert. Sie erfolgt aus einer weiteren statistischen Kennzahl, nämlich der Varianz (s^2).²² Die gängigen Computerprogramme MS-Excel, Open Office) enthalten den Berechnungsbefehl StAbw, so dass die gewünschten Ergebnisse unschwer zu erhalten sind.

Dazu das Beispiel aus einer Studiengruppe zur Bewertung der Praxisorientierung eines Lehrenden:

25 Studierende antworten wie folgt:

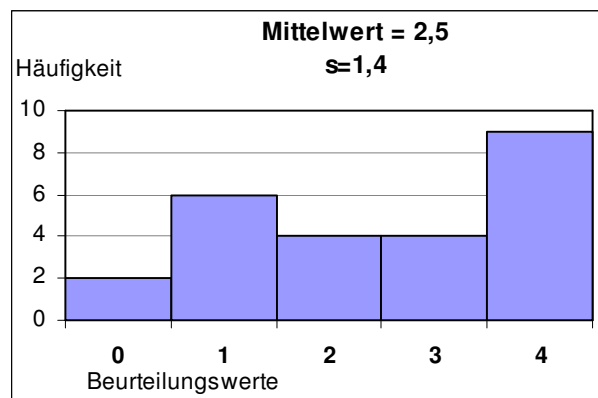
2 x trifft gar nicht zu = 2 x 0

6 x trifft eher nicht zu = 6 x 1

4 x teils – teils = 4 x 2

4 x trifft eher zu = 4 x 3

9 x trifft voll zu = 9 x 4



Der Mittelwert wäre 2,5 (gerundet auf eine Stelle nach dem Komma, um eine Scheingenaugkeit zu vermeiden). Der Lehrende könnte von einem zufriedenstellenden Ergebnis ausgehen. Dass mehr als ein Viertel der Studierenden der Auffassung ist, dass die Lehrveranstaltung nicht oder nur wenig praxisorientiert ist, wird aus dem Mittelwert nicht deutlich. Erst wenn dazu die errechnete Standardabweichung von 1,4 als Information geliefert wird, wird deutlich, dass der

²¹ Andere Streuungsmaße sind die **Spannweite** = Breite des Spektrums von der kleinsten zur größten Ausprägung und der **Quartilsabstand** = Differenz zwischen dem ersten und dem dritten Quartil einer Verteilung.

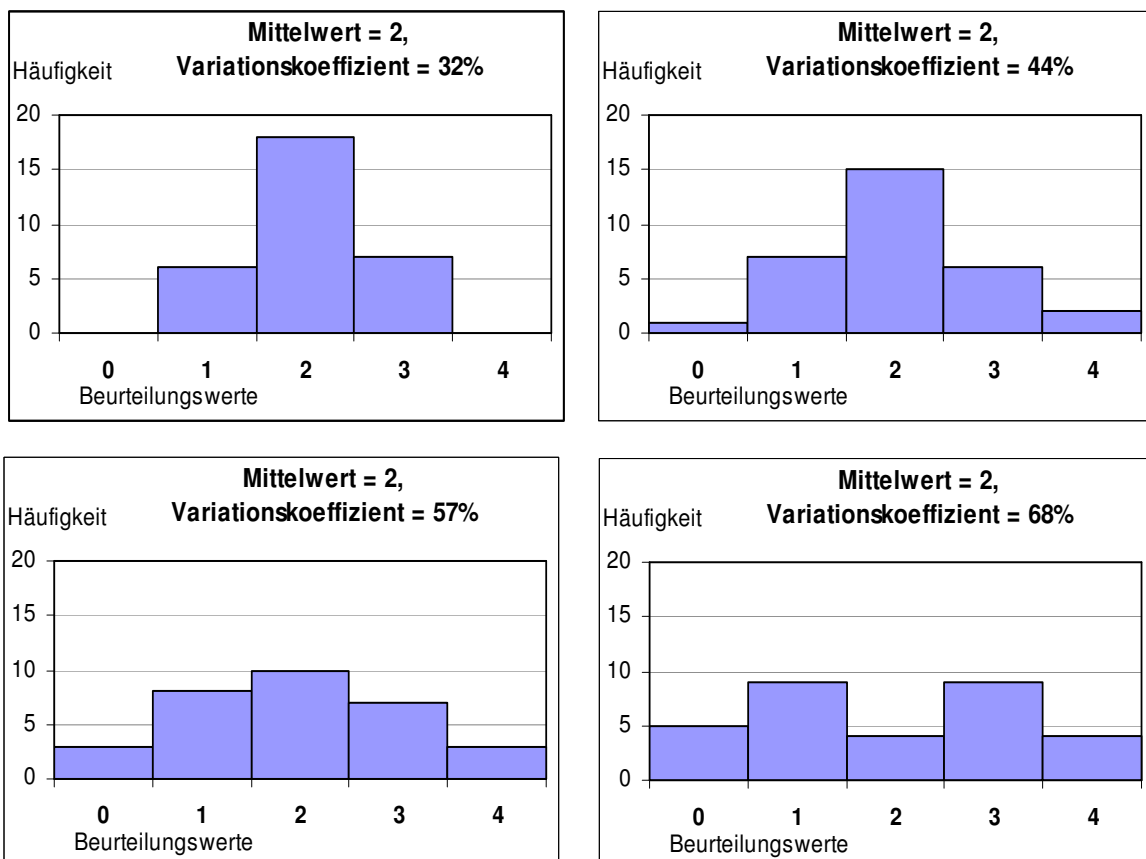
²² Die Varianz ist die durchschnittliche quadrierte Abweichung der Werte vom Mittelwert. Abweichungen der einzelnen Werte vom Mittelwert werden durch Quadrierung verschieden stark gewichtet. Berechnung: Summe der quadrierten Abweichungen der Werte vom Mittelwert, dividiert durch die Anzahl aller Werte. Die Wurzel aus der Varianz – um eine Maßzahl mit der gleichen Einheit wie die ursprünglichen Werte zu erhalten – ist die Standardabweichung.

Mittelwert nur bedingt hilfreich ist. Dazu muss der Lehrende auch wissen, dass eine Standardabweichung von über 1 meist ein Indikator für eine hohe Unterschiedlichkeit der Ergebnisse ist. In der grafischen Darstellung oben wird dieses Ergebnis erkennbar.

Sollte eine Standardabweichung einmal den Wert 0 erreichen, bedeutet dies, dass alle Befragten dieselbe Bewertung abgegeben haben (z. B. 25 x trifft voll zu).

Als relatives Streuungsmaß kann auch der *Variationskoeffizient* ausgewiesen werden. Er wird als Quotient aus Standardabweichung und arithmetischem Mittel berechnet und beschreibt den prozentualen Anteil der Standardabweichung am arithmetischen Mittel.²³ Die Berechnung: Standardabweichung geteilt durch Mittelwert, multipliziert mit 100. Die folgende Abbildung zeigt bei einem Mittelwert von 2 mögliche unterschiedliche Verteilungen und Variationskoeffizienten.

Abbildung 1 Streuungen um den Mittelwert



²³ Varkoeff. = Std.abw. / Mittelwert * 100.

Mit dem Variationskoeffizienten wird auch ein Vergleich der Streuungen bei unterschiedlichen Mittelwerten möglich. Bei Variationskoeffizienten von über 66% sind die Einschätzungen immer sehr unterschiedlich und das arithmetische Mittel kann die Verteilung nicht repräsentieren. In dem Beispiel oben (Mittelwert=2,5, Standardabweichung=1,4) liegt der Variationskoeffizient bei 56%.

5.4 Zustimmungswerte (Ja%-Werte) als verdichteter Kennwert

Der sogenannte Ja%-Wert ist wie der Mittelwert eine verdichtete Kennzahl. Hier wird nur die Anzahl der Personen – in Prozent ausgedrückt – genannt, die einer Aussage zugestimmt hat. Der Mittelwert ist wie oben beschrieben nämlich von der zugrundeliegenden Kodierung und Polung der Skala abhängig und damit nicht selbsterklärend oder für jeden transparent.²⁴ Die Ja%-Werte haben Vorteile bei der Kommunikation der Daten, weil Nachfragen nach der Polung der Skala und Zahl der Antwortkategorien entfallen. Ein Prozentwert ist jedem vertraut, so dass auch eine grafische Darstellung entbehrlich ist.

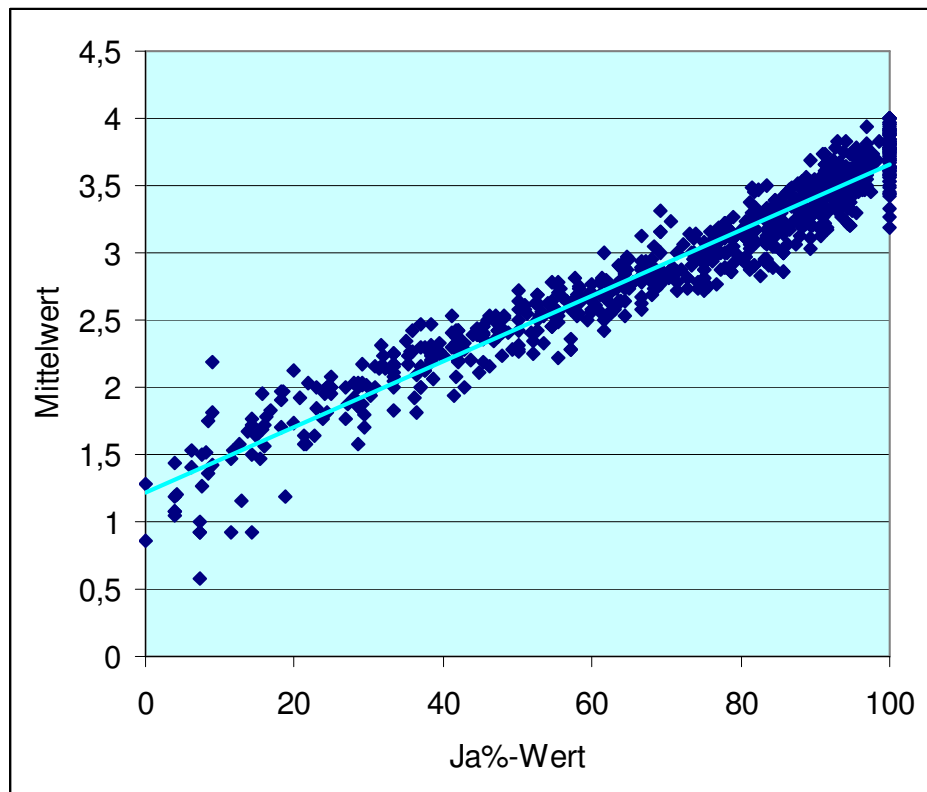
Die Ja%-Statistik kennzeichnet die Anzahl von Personen, die die Kategorie „trifft eher zu“ oder „trifft voll zu“ angekreuzt hat im Verhältnis zur Gesamtzahl der Befragten.

Allerdings beschreibt der Ja%-Wert die zentrale Tendenz der Bewertung nicht so genau wie der Mittelwert, da die Intensität der zustimmenden Antworten nicht berücksichtigt wird und „teils/teils“ oder ablehnende Antworten ignoriert werden. Der Mittelwert berücksichtigt diese Informationen, weil er – bei der hier verwendeten 5erSkala – die Besetzung aller fünf Kategorien repräsentiert. Mittelwert und Ja%-Statistik korrelieren jedoch hoch positiv.²⁵ Das heißt: je höher der Mittelwert ist, desto höher ist der Zustimmungswert, im Ja%-Wert ausgedrückt. Die folgende Abbildung zeigt die empirische Beziehung von Ja%-Wert zum Mittelwert am Beispiel der durchschnittlichen Bewertungen der Lehrveranstaltungen ihrer Dozenten, die die Studierenden in einer Untersuchung an der Verwaltungsfachhochschule in Wiesbaden abgegeben haben.

²⁴ Die Frage, ob ein Mittelwert von 1,7 gut oder schlecht ist, kann man nur beantworten, wenn die Polung der Skala (0 bedeutet „trifft gar nicht zu“ oder 1 bedeutet „trifft genau zu“), die Kodierung (z. B. von 0 bis 4 oder von 1 bis 5) und die Anzahl von Antwortstufen (5-stufig) bekannt sind.

²⁵ Siehe Abbildung 2: Korrelationskoeffizient $r=0,96$.

Abbildung 2 Zusammenhang von Ja%-Werten zum Mittelwert



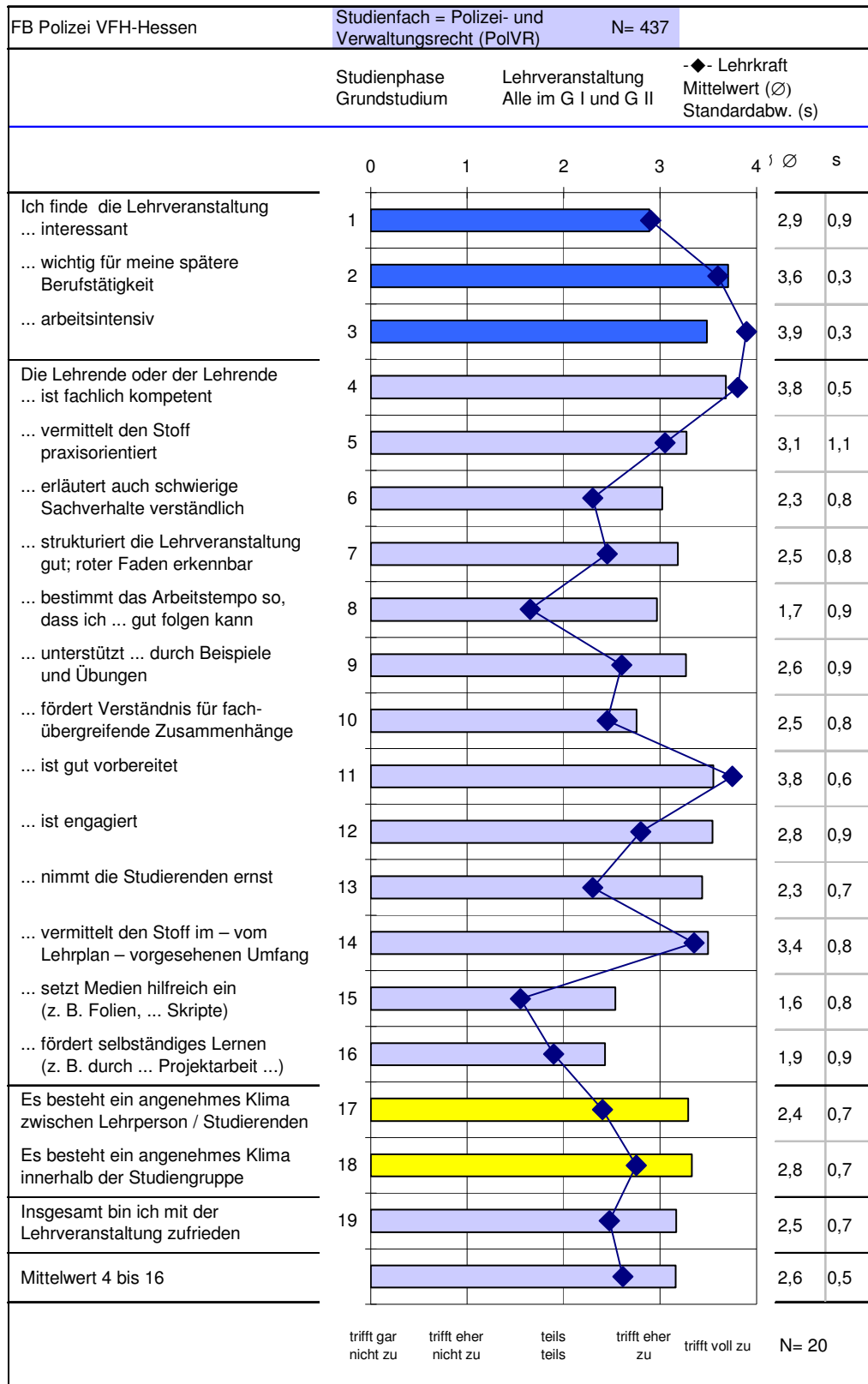
5.5 Darstellung von Ergebnissen

Die Expertengruppe Evaluation und Qualität hat 2004 eine Darstellung dozentenbezogener Ergebnisse durch ein Profil entwickelt.²⁶ Dieses Profil ist eine individuelle Rückmeldung für eine oder einen Lehrenden; es kann erstellt werden, sofern mehr als 10 Studierende eine Bewertung abgegeben hatten.

Die Abbildung 3 zeigt das Beispiel eines Profiles. Es handelt sich um ein kombiniertes Säulen-Liniendiagramm. Für jedes einzelne Item (so wird die zu bewertende Aussage genannt) kann die oder der Lehrende eigene durchschnittliche Ergebnisse (Punkt, mit Linie verbunden) mit den durchschnittlichen Ergebnissen aller Lehrenden (haupt- und nebenamtliche) in dem Fach vergleichen, zu dem sie oder er gehört. Überdurchschnittliche, durchschnittliche oder unterdurchschnittliche Werte sind in dieser Form gut erkennbar.

²⁶ Vgl. im Folgenden Gesamtbericht der Expertengruppe Evaluation 2004. Online-Dokument: http://www.verwaltungsmanagement.info/eva/2004/Gesamtbericht_Exp_Evaluation_2004.pdf

Abbildung 3 Auswertungsprofil – Beispiel aus dem Fach Polizei- und Verwaltungsrecht (PoIVR)²⁷



²⁷ Beispiel aus dem Qualitätsbericht 2005 der Verwaltungsfachhochschule in Wiesbaden.

Unter den Aussage-Items des Fragebogens wird ein Mittelwert über die dozentenbezogenen Items 4 bis 16 – die die eigentliche Lehre betreffen – ausgegeben²⁸. Die Datentabelle neben der Grafik enthält den Mittelwert der Lehrkraft sowie die Standardabweichung, die den Lehrenden über die Unterschiedlichkeit der Antworten der Verteilung informiert.

Für die Erstellung eines Profils steht eine Exceldatei zur Verfügung. Diese Art von Profilen und individueller Rückmeldung wurde nach dem Pilotprojekt an die Lehrenden verschickt und wird in den Fachhochschulen im Rahmen von Qualitätsberichten weiterhin verwendet.

Für die Interpretation der dozentenspezifischen Ergebnisse zu den verschiedenen Aspekten der Lehre können „Grenzwerte“ als Maß herangezogen werden, sinnvoll allerdings nur bezogen auf die verschiedenen Fachgebiete. Denn je nach Fachgebiet werden die Lehrenden unterschiedlich beurteilt. An dieser Stelle sollen die Ergebnisse des Pilotprojektes der Experten 2004 zitiert werden:

„Die Bewertung der Lehrenden durch die Studierenden ist unter anderem davon abhängig, ob das jeweilige Fach die Studierenden interessiert und – allerdings in geringerem Maße – ob die Relevanz eines Faches für die spätere Berufstätigkeit hoch eingeschätzt wird. Mit anderen Worten: Es kann nicht davon ausgegangen werden, dass im Rahmen einer studentischen Lehrveranstaltungskritik personenbezogene Merkmale isoliert bewertet werden. Vielmehr muss der Gesamtkontext, in den Fächer und Veranstaltungen eingeordnet werden, als wichtige Determinante für die Urteile der Studierenden berücksichtigt werden.

Die Möglichkeit einer Heranziehung von Daten aus der Lehrevaluation von Studierenden zu Zwecken der Personalbeurteilung der Lehrenden sind damit stark eingeschränkt. Eine vergleichende Beurteilung von Lehrenden ist bestenfalls innerhalb bestimmter Fächergruppen bzw. nur innerhalb bestimmter Einzelfächer möglich.“ (S. 47)

Die empirischen Ergebnisse aus sieben Fachhochschulen, die an dem Pilotprojekt 2003/2004 teilgenommen haben, zeigt die folgende Tabelle:

²⁸ Ein Mittelwert über alle Fragen erlaubt keine Aussage über die Einschätzung der Lehre, es wäre ein methodischer Fehler, diesen Mittelwert z. B. bei der Vergabe von Leistungszulagen zu verwenden.

Tabelle 1 Mittelwerte in den Fachgebieten
(ca. 35.000 Datensätze aus 7 Fachhochschulen) 2004

Fachgebiet	Fachbereich Verwaltung		Fachbereich Polizei	
	N	Mittelwert	N	Mittelwert
Rechtswissenschaft	5574	3,0	6353	2,9
Wirtschaftswissenschaft	2317	2,9	1013	2,7
Sozialwissenschaft	1579	2,6	2679	3,0
Polizeiwissenschaft			5435	3,1

5.6 Benchmarking

Wenn Hochschulen gleiche Fragebogen verwenden und nach denselben Kriterien auswerten, lassen sich so wie in dem Pilotprojekt Vergleiche über Landesgrenzen hinweg anstellen und ein Benchmarking vornehmen.

Die Untersuchung der Experten 2004 kann als Vergleichsmaßstab in den Fachhochschulen verwendet werden. Ein Beispiel dafür zeigt die folgende Tabelle.²⁹ Dargestellt werden die Ergebnisse der Studie der Experten Evaluation und Qualität im Vergleich zu Ergebnissen der Verwaltungsfachhochschule in Wiesbaden (VFH). 2003/2004 wurden Daten von drei verschiedenen Fachbereichen Verwaltung (einschließlich des FB Verwaltung der VFH Hessen) und fünf verschiedenen Fachbereichen Polizei (ohne den Fachbereich Polizei der Verwaltungsfachhochschule; diese Evaluation fand erst ein Jahr später statt) ausgewertet.

Tabelle 2 Vergleich der Ergebnisse – FB Polizei und FB Verwaltung 2003/2004
(Vergleich mit anderen Fachhochschulen) und 2009 VFH Hessen

Bewertungsaspekte	Fachbereich Polizei				Fachbereich Verwaltung			
	2003/2004 (ohne VFH)		2009 VFH		2003/2004		2009	
	Mittelwert	Std. abw.	Mittelwert	Std. abw.	Mittelwert	Std. abw.	Mittelwert	Std. abw.
Gesamturteil	3,0	1,0	3,0	1,0	2,8	0,9	2,7	1,0
Mittelwert errechnet (4-16)	3,0	0,6	3,1	0,7	2,9	0,6	2,9	0,8

5 FH, N=16.793 Hessen, N=923 3 FH, N=10.473 Hessen, N=836

Der Fachbereich Polizei der VFH wird 2009 im Vergleich zu anderen Fachhochschulen mit dieser Studienausrichtung gleich oder – bei den dozentenspe-

²⁹ Quelle: Qualitätsbericht der Verwaltungsfachhochschule in Wiesbaden, 2009.

zifischen Werten – geringfügig besser bewertet. Im Fachbereich Verwaltung der VFH wird die Lehre im Gesamturteil – beim Vergleich zu den zusammengefassten Ergebnissen 2003 – etwas schlechter beurteilt.

Tabelle 3 zeigt die Ergebnisse aus 2003/2004 für die einzelnen Fachbereiche Polizei der FHöD im Vergleich zu den Ergebnisse der VFH, FB Polizei:

Tabelle 3 Vergleich der Ergebnisse – FB Polizei 2003/2004 (Vergleich mit anderen Fachhochschulen) und Ergebnisse VFH Hessen 2004 und 2009

	Schleswig-Holstein	NRW (Abt. Köln)	Bremen	Sachsen-Anhalt	Berlin	Hessen	Hessen
Bewertungsaspekte	2003/2004					2004	2009
	Mittelwert						
Gesamturteil	2,8	3,2	2,8	3,4	2,7	3,2	3,0
Mittelwert errechnet (4-16)	3,0	3,2	3,0	3,0	2,9	3,2	3,1
	N=5.232	N=5.041	N=3.300	N=1.228	N=1.474	N=5.258	N=923

Im FB Polizei in Hessen wurden 2004 und 2009 überdurchschnittlich gute Ergebnisse erreicht. In NRW (Abt. Köln) wurden ähnlich gute Ergebnisse 2003/2004 erreicht.

Schließlich der Vergleich im Fachbereich Verwaltung.

Tabelle 4 Vergleich der Ergebnisse – FB Verwaltung 2003/2004 (Vergleich mit anderen Fachhochschulen) und VFH Hessen 2009

Bewertungsaspekte	Schleswig-Holstein		NRW (Abt. Köln)		Hessen		Hessen	
	2003/2004		2003/2004		2003/2004		2009	
	Mittelwert	Std. abw.	Mittelwert	Std. abw.	Mittelwert	Std. abw.	Mittelwert	Std. abw.
Gesamturteil	2,7	0,9	2,8	1,0	3,0	0,9	2,7	1,0
Mittelwert errechnet (4-16)	2,8	0,6	2,9	0,7	3,0	0,6	2,9	0,8
	N=3.025		N=3.720		N=3717		N=836	

Im Fachbereich Verwaltung werden die überdurchschnittlich guten Ergebnisse von 2003 nicht mehr erreicht. Im Vergleich mit den einzelnen Fachbereichen der anderen FHöD zeigt sich jedoch, dass die Werte nicht schlechter sind als die aus NRW und Schleswig-Holstein 2003.

6 Wissensmanagement

6.1 Ausgangslage

Die Frage, wie Wissensmanagement geleistet, d. h. das in den Mitgliedshochschulen verfügbare Wissen zum Thema Evaluation systematisch aufbereitet und auch für andere erschlossen werden kann, beschäftigt die Expertengruppe schon seit Jahren. Eine befriedigende Antwort kann nicht gegeben werden. Grund dafür sind Rahmenbedingungen: Es stehen keine nennenswerten Kapazitäten für ein eigenständiges Wissensmanagement zur Verfügung. Allerdings bedarf Wissensmanagement auch nicht besonderer zur Verfügung gestellter Kapazitäten sondern sollte vor allem aus der Praxis heraus entstehen.

Erfahrungen mit bisher verfügbaren Möglichkeiten, insbesondere den Lernplattformen, haben keine brauchbaren Ergebnisse gezeigt. Die Handhabung verfügbarer Lernplattformen ist zu umständlich, als dass damit ein als hilfreich empfundenen Instrument zur Verwaltung von Wissensbeständen gestaltet werden kann. Das gilt insbesondere für ILIAS, die derzeit als kostenlos verfügbare Lernplattform eingesetzt wird.³⁰

Wissensmanagement lebt von der Aktivität derjenigen, die in dem Bereich tätig sind – dort ohnehin Informationen produzieren, sammeln, aufbereiten. Ein möglicher Weg zur Verwirklichung eines solchen Wissensmanagements könnten Programme sein, die im Internet das Zusammentragen von Informationen und Gestalten von Informationsangeboten durch die Nutzerinnen und Nutzer organisieren, so genannte kollaborative Programme wie Foren, Blogs oder Wikis (mit der bekanntesten Anwendung „Wikipedia“).

Die Expertengruppe hat begonnen, die Materialien mit einem solchen Programm zu erfassen und bereitzustellen. Es handelt sich um eine Variante eines Wikis, einer Software, mit der die Nutzerinnen und Nutzer selbst Beiträge verfassen und bearbeiten können, wobei Fehler und Manipulationen durch die Speicherung aller Versionen korrigierbar bleiben. Auch gibt es eine Rechtever-

³⁰ Alternativen im Bereich der Fachhochschulen für den öffentlichen Dienst sind zumeist noch nicht erprobt worden, obwohl einige davon nach vorliegenden Erfahrungsberichten wesentlich benutzerfreundlicher sein sollen. Immerhin hat auch die Fernuniversität Hagen inzwischen die ebenfalls kostenlos verfügbare Lernplattform Moodle eingeführt, eine durch eine weltweite Anwendergemeinschaft entwickelte und weiter gepflegte Lernplattform, die deutlich benutzerfreundlicher ist als ILIAS. Gerade für kleinere Hochschulen, die sich nicht den Luxus einer eigenständigen Weiterentwicklung leisten können, könnte die neue, in der Grundform ebenfalls kostenlose Lernplattform Dokeos interessant sein, weil sie anders als ILIAS und Moodle bereits wesentlich besser vorstrukturierte Bausteine für Lernangebote bereit stellt.

waltung, die es erlaubt, den Zugriff insgesamt und auf einzelne Seiten abgestuft zu regeln.

Es zeigte sich, dass diese Arbeiten nur einen bescheidenen Erfolg hatten.

- Trotz aller Erleichterungen, die dieses Programm bereits bietet, erfordert es für eine insgesamt befriedigende Gestaltung der Beiträge doch immer noch einen Grad an IT-Kompetenz, den der normale Nutzer nicht hat.
- Das Erstellen und Pflegen von Beiträgen, ihrer Einordnung in eine logische Struktur und die Verlegung zu anderen Beiträgen erfordert einen Aufwand, der als Zusatzaufwand zu normalen Tagesarbeit nicht regelmäßig zu leisten ist, beziehungsweise der immer wieder die Unterstützung erfordert.

Anders ausgedrückt: die Programme sind immer noch nicht komfortabel genug. Schließlich gilt auch hier die 90-9-1-Regel, nach der bei solchen, auf die Mitarbeit der Nutzer/innen angewiesenen Programmen der überwiegende Teil passiv bleibt und nur ein ganz kleiner Prozentsatz wesentliche und umfangreiche Beiträge liefert.

6.2 Empfehlungen

Die Expertengruppe ist jedoch der Meinung, dass das Potenzial, das in den Informationsbeständen der Mitgliedshochschulen und in dem Wissen ihrer Mitglieder vorhanden ist, nicht ungenutzt bleiben sollte. Auch bei Nutzung komfortabelster Programme ist jedoch eine zentrale Betreuung des Systems und Unterstützung der Nutzer notwendig – also eine Person, die durch entsprechende Freistellung ihr Engagement auch einbringen kann, ohne sich dafür jeweils rechtfertigen zu müssen, weil andere dienstliche Aufgaben vernachlässigt werden oder dieses Engagement durch unzumutbaren Verzicht auf Freizeit erkauft werden muss.

Wissensmanagement im Bereich von Evaluation und Qualität setzt also eine Ansprechperson voraus, der dafür mindestens 10% einer Vollzeitstelle (vier Zeitstunden wöchentlich) zur Verfügung stehen. Es wird angeregt, diese Ressourcen bereitzustellen, die sich durch Einsparungen bei der Evaluationsarbeit und durch bessere Ergebnisse rechnen.³¹

Technisch erscheint eine Lösung optimal, die ein Wiki mit einem virtuellen Laufwerk kombiniert. Das virtuelle Laufwerk im Internet ist über Internet-

³¹ Bei Berechnung mit Teil- und nicht mit Vollkosten entspricht das einem Aufwand von jährlich etwa 6.000 € nach den Personalkostentabellen des BMF.

Adressen zugänglich und speichert Dokumente, die jederzeit unter gleichem Namen in aktualisierter Fassung hoch geladen werden können. Inhaltliche Beiträge und die Erschließung aller Wissensbestände erfolgen über ein Wiki, zu dem alle, die jeweils mit der Information zu tun haben, beitragen können.

Diese beiden Ressourcen sind heute für etwa 10 € monatlich zu bekommen, so dass Sachkosten in der Größenordnung von 120 € pro Jahr entstehen würden. Möglich wäre auch, dass bereits vorhandene Internet-Auftritte an einer Mitgliedshochschule diese Leistungen bereitstellen. Jedoch ist unbedingt erforderlich, dass eine ausreichende Leistungsfähigkeit gewährleistet ist. Es sollte deutlich weniger als 10 Sekunden dauern, bis das System reagiert.

Literaturverzeichnis

- Balk, Michael: Die Evaluation von Lehrveranstaltungen. Die Wirkung von Evaluationsrückmeldung. Frankfurt a. M. 2000
- Bonner Modell der Hochschulevaluation: http://de.wikipedia.org/wiki/Bonner_Modell_der_Hochschulevaluation (11.03.2010)
- Costa, René: Das Bonner Modell der Hochschulevaluation: Erste Erfahrungen mit der Implementierung einer webbasierten Modul- und Lehrveranstaltungsevaluation. In: Georg Rudinger / Katharina Hörsch / Thomas Krüger (Hrsg.): Forschung und Beratung – Das Zentrum für Evaluation und Methoden (S. 49-54). Göttingen 2009
- Dresel, Markus / Rindermann, Heiner / Tinsner, Karen: Beratung von Lehrenden auf der Grundlage studentischer Veranstaltungsbeurteilungen. In: Kluge, Annette / Schüler, Kerstin (Hrsg.): Qualitätssicherung und -entwicklung an Hochschulen: Methoden und Ergebnisse (S. 193-204). Lengerich 2007
- Expertengruppe Evaluation und Qualität: Pilotprojekt „Studierendenbefragung“ 2004. Online-Quelle: http://www.verwaltungsmanagement.info/eva/2004/Gesamtbericht_Exp_Evaluation_2004.pdf
- Fredersdorf, Frederic / Lehner, Martin: Hochschuldidaktik und Lerntransfer. Bildungscontrolling von FH-Studiengängen. Bielefeld 2004
- Hochschulrektorenkonferenz (Hrsg.): Qualitätsentwicklung an Hochschulen. Erfahrungen und Lehren aus 10 Jahren Evaluation. Dokumentation zur gleichnamigen Veranstaltung des Projekts Qualitätssicherung der HRK am 3. und 4. November 2005 im Bonner Wissenschaftszentrum. Beiträge zur Hochschulpolitik 8/2006. Online-Quelle: http://www.hrk.de/de/projekte_und_initiativen/4350.php (10.03.2010)
- KMK: Qualitätssicherung der Lehre. Beschluss der Kultusministerkonferenz vom 22.09.2005. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_09_22-Qualitaetssicherung-Lehre.pdf (10.03.2010)
- Krawietz, Marian (HIS Hochschul-Informationen-System): Evaluation der Evaluationen – Die Perspektive der Studierenden. 31.10-2006 (evaNet-Position 08/2006), http://www.hrk-bologna.de/de/projekte_und_initiativen/4160.php (11.03.2010)
- Kromrey, Helmut: Qualität und Evaluation im System Hochschule. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. 2. Aufl., Opladen 2003,

S. 233-258. Online-Quelle: http://www.profkromrey.de/Kromrey_Eval_Univ.pdf (11.03.2010)

Kuckartz, Udo / Ebert, Thomas / Rädiker, Stefan / Stefer, Claus: Evaluation Online. Internetgestützte Befragung in der Praxis. Wiesbaden 2009

Meyer, Helga / Frank, Gudrun / Janas, Dana: Kompetenzerwerb im Studium messen. Das Projekt „moreQ“ der Hochschule Bremen. In: Wissensschaftsmanagement 4, Juli/August 2007, S. 25-32. Online-Quelle: http://www.lemmens.de/verlag/zeitschriften/wima_archiv/wima_2007/wima_4_07.pdf (10.03.2010)

Nickel, Sigrun: Qualitätsmanagementsysteme an Universitäten und Fachhochschulen: Ein kritischer Überblick. Beiträge zur Hochschulforschung, Heft 1, 30. Jahrgang, 2008. Online-Quelle: http://www.che.de/downloads/Nickel_QM_Ueberblick.pdf (10.03.2010)

Rindermann, Heiner / Kohler, Jürgen / Meisenberg, Gerhard: Quality of Instruction Improved by Evaluation and Consultation of Instructors. In: International Journal for Academic Development, Vol. 12, No. 2, November 2007, pp. 73-85. Online-Quelle: <http://www.uni-graz.at/pslgcwww/rindermann/publikationen/07IntJourAD.pdf> (12.02.2010)

Schaeper, Hilde / Briedis, Kolja: Kompetenzen von Hochschulabsolventinnen und Hochschulabsolventen, berufliche Anforderungen und Folgerungen für die Hochschulreform. Projektbericht (HIS), 2004. Online-Quelle: http://www.bmbf.de/pub/his_projektbericht_08_04.pdf (12.02.2010)

Von der Lippe, Peter / Kladroba, Andreas: Repräsentativität von Stichproben. In Marketing ZFP 24, 2002, S. 139 - 145. Online-Quelle: <http://von-der-lippe.org/dokumente/Repraesentativitaet.pdf> (10.08.2010).

Wissenschaftsrat: Empfehlungen zur Qualitätsverbesserung von Lehre und Studium. Drs. 8639-08. 2008. Online-Quelle: http://www.exzellentelehre.de/pdf/empfehlungen_zur_qualitaetsverbesserung_von_lehre_und_studium_2008.pdf (12.02.2010)

Zumbach, Jörg / Spinath, Birgit / Schahn, Joachim / Friedrich, Monika / Kögel, Melanie: Entwicklung einer Kurzsкала zur Lehrevaluation. In: Michael Krämer / Siegfried Preiser / Kerstin Brusdeylins (Hrsg.): Psychologiedidaktik und Evaluation Bd. VI, Göttingen 2007, S. 317-325. Online-Quelle: http://www.psychologie.uni-heidelberg.de/ae/diff/js/evaluation_kurzskala.pdf (23.02.2010)

Weitere Quellen und zusätzliche Angaben enthält das ständig aktualisierte, als Wiki gestaltete Verzeichnis im Internet:

<http://olevde.wiki.zoho.com/Quellen-zur-Evaluation-in-Hochschulen.html>

Anhang: Empfehlungen des Wissenschaftsrates 2008 (Auszug)

Auszug aus den Empfehlungen des Wissenschaftsrates zur Qualitätsverbesserung von Lehre und Studium, Drs. 8639-08, 2008

Ausrichtung der Hochschullehre (S. 8)

Die Studierenden in ihrem Lernen bestmöglich zu unterstützen, steht im Mittelpunkt aller Anstrengungen der Lehrenden und der Hochschulen in Studium und Lehre. Lehrende sollten selbstorganisiertes Lernen fördern und die Studienprozesse auf die Aneignung von fachlichen sowie überfachlichen Kompetenzen ausrichten. Eigeninitiative und Eigenverantwortung der Studierenden sollten gleichermaßen gefördert und eingefordert werden. Ein solches Rollenverständnis von Lehrenden und Lernenden ist zu flankieren mit Veranstaltungsformen, die förderliche Lehr- und Lernsituationen schaffen und das aktive Lernen unterstützen.

Qualitätsmanagement und Personalentwicklung (S. 10 f.)

Eine der zentralen Voraussetzungen für eine gezielte und kontrollierte Qualitätsentwicklung im Bereich Lehre und Studium ist der Aufbau eines systematischen und integrierten hochschulinternen Qualitätsmanagements. Der Wissenschaftsrat empfiehlt, dass alle Hochschulen innerhalb der nächsten drei bis fünf Jahre ein internationalen Standards entsprechendes Qualitätsmanagement aufbauen sollten. Dazu gehört es, im Zuge der Profilbildung ambitionierte Ziele festzulegen, Verantwortlichkeiten klar zuzuordnen und Anreize für ein verstärktes Engagement in der Lehre zu setzen. Das unter möglichst breiter Beteiligung aller Statusgruppen organisierte Qualitätsmanagement sollte als strategisches Steuerungsinstrument der Hochschulleitung die Stetigkeit eines definierten Leistungsniveaus sichern, Veränderungsprozesse fördern und mögliche Fehlentwicklungen frühzeitig identifizieren helfen.

Leistungstransparenz und Qualitätsbewertung (S. 11)

Der Wissenschaftsrat empfiehlt, im gesamten Hochschulsystem eine umfassende Praxis der Rechenschaftslegung und der Transparenz der Leistungsfähigkeit im Bereich Studium und Lehre zu entwickeln und zu etablieren. Dies stellt eine Grundvoraussetzung einer gezielten qualitätsorientierten Binnensteuerung der Hochschulen dar. **Alle Bewertungsverfahren sollten außer den erbrachten Leistungen (Output) systematisch auch die jeweiligen Voraussetzungen (Input) berücksichtigen.** (Hervorhebung ergänzt)

Zentrale Empfehlungen, Abschnitt B.III.: Qualitätsbewertung von Studium und Lehre (S. 82, Umrahmung im Original)

- Aufbau verlässlicher Bewertungsinstrumente für die Qualität der Lehrleistungen und zur differenzierten Erfassung des Kompetenzerwerbs im Studium.
- Berücksichtigung des Verhältnisses zwischen Input- und Output-Faktoren in Bewertungsverfahren der Qualität von Lehre und Studium.
- Herstellung von Transparenz über zentrale Indikatoren, die Rückschlüsse auf die Qualität von Studienangeboten zulassen.