

Rezension zu:

Großmann, Daniel/Wolbring, Tobias (Hg.): *Evaluation von Studium und Lehre. Grundlagen, methodische Herausforderungen und Lösungsansätze*. Wiesbaden: Springer VS, 2016. 429 Seiten, 34,99 EUR, ISBN 978-3-658-10885-4

*Antonia Baumeister*¹

Der von Daniel Großmann und Tobias Wolbring herausgegebene Sammelband „Evaluation von Studium und Lehre“ stellt Grundlagenwissen und neue Forschungsergebnisse zu methodischen und inhaltlichen Fragestellungen der Evaluation an Hochschulen dar. Die 21 Autor(inn)en des Bandes stammen aus verschiedenen Fächern und verdeutlichen bereits die Interdisziplinarität der Thematik. Der Band wird mit einem Geleitwort des Doyens der deutschen Hochschulevaluationsforschung, Hans-Dieter Daniel von der ETH Zürich, eingeleitet. Er skizziert die Entwicklung der Evaluation von Studium und Lehre in Deutschland von der Novelle des Hochschulrahmengesetzes im Jahr 1998 bis heute. Betont wird der hohe wissenschaftliche Anspruch von Evaluationen, der sich auch in den Beiträgen des Bandes widerspiegelt. Der Band gliedert sich in „Grundlagen“, „Input“, „Prozess“ und „Output“.

Grundlagen

Begonnen wird mit einem Überblick zu „Stand und Herausforderungen der Evaluation an deutschen Hochschulen“ (Daniel Großmann und Tobias Wolbring).

Der zweite Beitrag „Entstehung und Grundlagen der Evaluation“ von Reinhard Stockmann äußert interessantes interkulturelles Expertenwissen zur historischen Entwicklung der Evaluation in den USA versus Deutschland. Ausgehend von einem Forschungsmodell der Evaluation lässt Reinhard Stockmann auch kritische Töne an der deutschen Praxis der unabhängigen Evaluation eines Politikfeldes anklingen (vgl. S. 33-34).

Der dritte Beitrag „Evaluation, Kausalität und Validität“ von Tobias Wolbring ist metho-

disch grundlegend und unterstützt das Verständnis über veranschaulichende Beispiele. Manche fachspezifischen methodischen Begriffe, insbesondere zu statistischen Verfahren, z.B. ‚Sensitivitätsanalyse‘ oder ‚Stratifikation‘, hätte man noch näher erläutern können. Vielleicht könnte eine zweite Auflage von einem Glossar am Buchende profitieren. Intensiv setzt sich Wolbring mit der Bologna-Reform und ihrer Evaluation auseinander. Kritisch summiert er hier, dass noch keine „belastbare evidenzbasierte Wirkungsevaluation der Bologna-Reform“ (S. 83) erfolgt sei. Wenngleich die Forderung nach einer solchen Wirkungsevaluation aus wissenschaftlicher Sicht richtig ist, erscheint es praktisch kaum umsetzbar, internationale Reformen in einer für diesen Rahmen angemessenen Feldstudie zu untersuchen. Andere Bemühungen, Bildungssysteme vergleichend zu evaluieren (z.B. PISA-Studien), kranken an einer Reihe von methodischen Problemen (vgl. Rindermann/Baumeister 2015). Ein Verbund von Hochschulen zu einer Arbeitsgruppe könnte womöglich einen Ex-post-facto-Vergleich von Diplom- und Masterstudiengängen realisieren.

Input

Im ersten Beitrag des Input-Abschnitts „Soziale Determinanten der Studienaufnahme und Fachwahl“ von Werner Georg und Tino Bargel werden, basierend auf dem Deutschen Studierendensurvey, die teilweise sehr deutlichen Zusammenhänge zwischen der beruflichen Fachrichtung der Väter und der Fachrichtung der Studierenden aufgezeigt. Aufschlussreich sind auch die Ergebnisse zum Einfluss des Geschlechts auf die Fachwahl sowie die Ergebnisse zu fachkulturellen Distanzen. Folgerichtig spre-

1 Technische Universität Chemnitz

chen sich die Autoren bei Evaluationen für eine stärkere Berücksichtigung sozialer Merkmale der Studierenden aus.

Daniel Großmann berichtet in seinem Beitrag („Studienmotivation und Evaluation“) von einer Studie zu den Zusammenhängen zwischen Studienmotivation und Evaluation in sozialwissenschaftlichen Masterstudiengängen an der Universität Leipzig. Das Konstrukt der Studienmotivation wird ausführlich definiert. Dabei wird ein Dreikomponentenmodell zur Systematisierung herangezogen. Auf dieser relativ abstrakten Ebene bleibt es schwierig, das Zusammenspiel der drei Komponenten nachzuvollziehen und ihre Bedeutung und Funktion zu durchdringen; eine veranschaulichende Abbildung wäre hilfreich gewesen. Dennoch ist der Beitrag für alle interessant, die sich genauer über das Konstrukt der Studienmotivation, seine Messung sowie seinen Einfluss auf Evaluationen von Studium und Lehre informieren möchten. Der Beitrag ist mit 61 Seiten deutlich länger als die anderen Beiträge des Bandes. Insofern entstand beim Lesen der Eindruck eines ‚Buchens im Buch‘.

Methodisch ebenfalls exzellent ist die Studie „Messung von studentischem Workload“, von der Roger Berger und Bastian Baumeister berichten. In der Studie wurden im Wesentlichen zwei Erfassungsarten sowie zwei längsschnittliche Zeitperioden verglichen: Papiertagebuch versus Smartphone-App-Tagebuch, prospektiv versus retrospektiv, wobei die Retrospektiverhebung zur Erfassung soziodemographischer Variablen um einen Online-Fragebogen ergänzt war. Die Studie ist sehr gut nachvollziehbar und anschaulich beschrieben. Sie liefert aussagekräftige Befunde sowie wertvolle Anregungen zur Messung des studentischen Workloads.

Prozess

Der erste Beitrag des Prozessabschnitts „Lehrveranstaltungsevaluation an Hochschulen“ stammt von Heiner Rindermann und widmet sich dem Einfluss von Rahmenbedingungen auf Qualität von Lehre und Ergebnisse von Lehrveranstaltungsevaluation. Die zentralen Konzepte der Lehrveranstaltungsevaluation werden erläutert. Studienergebnisse verdeutlichen die Grenzen von Evaluation zur Verbesserung der Lehrqualität. In den Unterkapiteln 4 und 5 stellt Rindermann auf eindrückliche und gleichsam amüsante Art und Weise dar, welche Bedeutung Rahmenbedingungen für die Qualität der Hochschulbildung haben. Ein unbequemer Beitrag, der die Missstände an deutschen Universitäten, auch im internationalen Vergleich, deutlich macht.

Im zweiten Prozessbeitrag „Konstruktvaliditätsprobleme von Lehrvaluationen“ zeigen Brigitte Metje und Udo Kelle auf, wie entscheidend *Think-Aloud-* oder *Probing-*Verfahren – als zweite Säule neben quantitativen Methoden der Testkonstruktion – sind, um die Konstruktvalidität von Lehrvaluationsinstrumenten zu prüfen. Auch dieser Beitrag ergänzt die vorherigen um grundlegende Definitionen. Ein empirisches Forschungsprojekt an der Universität Marburg wird vorgestellt, in dessen Rahmen Beantwortungsfehler und Verständnisdivergenzen bei Lehrveranstaltungsevaluationsinstrumenten mithilfe von kognitiven Interviews mit den Beantwortenden aufgedeckt wurden. Die dadurch deutlich eingeschränkte Konstruktvalidität der untersuchten Items wäre mit rein quantitativen Methoden nicht erkannt worden, denn diese Analysen fielen unauffällig aus. Dieser Beitrag ist also insbesondere für die Entwickler(innen) und Anwender(innen) von Lehrvaluationsinstrumenten relevant.

Der dritte Beitrag „Studierendenbefragungen als Panelstudie“ stammt von Philipp Pohlenz, Frank Niedermeier, Melinda Erdmann und Juliana Schneider. Die Autor(inn)en erläutern, wie Panelstudien als Längsschnittdaten genutzt werden können, um die Qualitätssicherung an Hochschulen auf eine kausal-empirisch belastbare Basis zu stützen. Studiengänge zielen darauf ab, fachliche und überfachliche Kompetenzen zu vermitteln. Panelstudien tragen u.a. dazu bei, sich an eine kausale Effektivitätsprüfung von Studiengängen anzunähern. Die Potsdamer Befragungen werden jährlich durchgeführt bis zur Exmatrikulation. Danach folgen mehrere Absolventen- und Exmatrikuliertenbefragungen. Die Befragungen liefern umfangreiche Informationen u.a. zu selbsteingeschätzten Kompetenzen, zur Studienmotivation, Studienzufriedenheit, zu wahrgenommenen Schwierigkeiten sowie zu Abbruch- bzw. Wechselgründen. Interessant ist beispielsweise, welche Gemeinsamkeiten und Unterschiede in der Veränderung der Aspekte der Studienmotivation zwischen den untersuchten Studiengängen beobachtet werden können. Aus den Verlaufsmessungen zur Studienmotivation werden sinnvolle Schlussfolgerungen zur Weiterentwicklung der Studiengänge abgeleitet. Die Autor(inn)en weisen darauf hin, dass die Aussagekraft dieser Panelstudie bislang dadurch eingeschränkt war, dass lediglich zwei Messzeitpunkte vorlagen, die jedoch künftig vermehrt werden sollen. Daniel Großmann berichtet in seinem Input-Beitrag von einer Studie zur Erfassung der Studienmotivation, wenn gleich es sich im Gegensatz zu der Panelstudie

von Pohlenz und Kolleg(inn)en bei Großmann um eine Querschnittsstudie handelte. Interessant wäre die Frage, ob beide Studien ähnliche Befunde lieferten. Beispielsweise kommen die Studien insofern zu unterschiedlichen Ergebnissen, als Großmann von einer, im Vergleich zu den Fachstudiengängen stärkeren, Unzufriedenheit von Lehramtsstudierenden berichtet, während die Potsdamer Panelstudie signifikante Anstiege bestimmter Motivationsaspekte bei Lehramtsstudierenden nachwies. Die Verlaufsabbildungen zu diesen Panelbefunden auf Seite 312 lassen vermuten, dass bei der Gruppe der Lehramtsstudierenden ein vergleichsweise geringer Anstieg der Studienmotivation vorliegt. Dies käme der Interpretation der Befunde von Großmann wiederum nahe (vgl. S. 161: „Sonderfall Lehramtsstudium“). Weniger in methodischer Hinsicht, sondern im Hinblick auf die Interessen von Leser(inne)n, sei kritisch angemerkt, dass die Leipziger Querschnittsstudie und die Potsdamer Panelstudie sowie ihre Befunde zur Studienmotivation nicht auf eine Vergleichbarkeit hin ausgelegt waren und es daher möglicherweise interessanter gewesen wäre, wenn die Befunde der Panelstudie zu den selbsteingeschätzten Kompetenzen der Studierenden anstatt zu ihrer Studienmotivation berichtet worden wären.

Output

Der erste Beitrag des Output-Abschnitts „Studienabbruchquoten als Evaluationskriterium“ stammt von Daniel Klein und Volker Stocké. Dieser Beitrag dürfte für zentrale hochschulpolitische Akteure, wie die Kultusministerkonferenz (KM), den Wissenschaftsrat (WR) sowie die Hochschulrektorenkonferenz (HRK) besonders informativ sein. Die Autoren kritisieren, dass Studienabbruchquoten häufig als Evaluationskriterium und Steuerungsinstrument der Qualitätssicherung genutzt werden, obwohl die Vergleichbarkeit und Validität der Messungen fragwürdig ist. Zudem weisen sie auf mögliche Risiken dieser Praxis im Hochschulbereich hin. Mit dem Monitoring des deutschen Hochschulsystems befassen sich z.B. die Autorengruppe Bildungsberichterstattung, das Statistische Bundesamt sowie das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung in Zusammenarbeit mit der Arbeitsgruppe Hochschulforschung der Universität Konstanz. Diese Institutionen kommen, aufgrund ihrer Berechnungsvarianten, zu deutlich unterschiedlichen Kennzahlen für die Studienabbruchquote. Beispielsweise unterscheidet

sich die Berücksichtigung von Bildungsinländer(inne)n und Bildungsausländer(inne)n. Des Weiteren diskutieren die Autoren die empirisch bislang mehr oder weniger einheitlich vorgefundenen Ursachen für Studienabbrüche, von denen manche außerhalb des Einflussbereichs der evaluierten Akteure liegen, z.B. soziodemographische Merkmale. Während die Befundlage zur Lehrqualität als Ursachenfaktor für Studienabbrüche uneinheitlich ist, hat sich die Note der Hochschulreifepfprüfung als wichtiger Ursachenfaktor für den Studienerfolg erwiesen. Einen Ansatz zur akademischen und sozialen Integration von Studierenden als Präventionsmaßnahme gegen Studienabbrüche halten die Autoren für schwer umsetzbar und wenig Erfolg versprechend.

Felix Wolter und Jürgen Schiener stellen in ihrem Beitrag zum Thema Kompetenzdiagnostik an der Hochschule „Methoden, Herausforderungen und aktuelle Entwicklungen“ (S. 367 ff.) vor. Dieser Beitrag ist für Evaluationsbeauftragte sowie für Forscher(innen) interessant. Es geht um die Frage der Vergleichbarkeit von Noten im Hochschulbereich angesichts u.a. dozentenpezifischer Gepflogenheiten. Die zentralen Konzepte und Prinzipien der Kompetenzdiagnostik, wie sie beispielsweise auch im Rahmen von Schulleistungsstudien angewandt werden, werden gut verständlich erläutert und u.a. bildlich dargestellt, beispielsweise Kompetenzstrukturmodell und Kompetenzniveaumodell, adaptives Testen, IRT-Verfahren usw. Die Bedeutsamkeit des „Differential Item Functioning“ (DIF; vgl. S. 377)² und seine Implikationen für den Hochschulbereich (vgl. z.B. S. 381) werden erklärt. Es werden Beispiele für deutsche und internationale Projekte zur Kompetenzmessung im Hochschulbereich genannt. Speziell für Deutschland wären dies eine Teilstudie des Nationalen Bildungspanels (NEPS) sowie der vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Forschungsverbund „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“. Methodische Anmerkungen sowie einige Ergebnisse des Pilotprojekts werden in aller Kürze vorgestellt und sind sehr interessant. Die Autoren nennen konkrete Anwendungsmöglichkeiten u.a. für die Studiengangsentwicklung.

Der abschließende Beitrag des Output-Abschnitts sowie des Sammelbandes, „Berufseinstiege von AkademikerInnen mit Migrationshintergrund in Österreich“, stammt von Roland Verwiebe und Melek Hacıoglu. Sie stellen ei-

2 Ein DIF liegt vor, wenn Test-Items nicht aufgrund der Personenfähigkeit, sondern aufgrund einer Drittvariablen Teilnehmergruppen diskriminieren.

nen Mixed-Methods-Ansatz vor, der über herkömmliche Absolventenbefragungen hinausgeht und dazu dient, Barrieren und Diskriminierungserfahrungen von Absolvent(inn)en zu identifizieren. An der Universität Wien wurden zeitgleich quantitative Fragebogendaten und qualitative Interviewdaten in einem ‚Convergent Parallel Design‘ erhoben. Die quantitativ untersuchten Faktoren umfassten beispielsweise die Dauer der Arbeitssuche beim Berufseinstieg, die Anzahl an Bewerbungen sowie die berufliche Stellung der ersten Tätigkeit. Interessanterweise wurden keine Lohnunterschiede zwischen Akademiker(inne)n mit versus ohne Migrationshintergrund bei ihrer ersten beruflichen Tätigkeit festgestellt. Die Probleme, die im Rahmen der Interviews genannt wurden, betrafen beispielsweise den ausländischen Namen oder Stereotype bezüglich des optimalen Aussehens, um für Kundengespräche einsetzbar zu sein. Die Autorin und der Autor fanden Hinweise darauf, dass sich die Benachteiligungen insbesondere bei Akademikerinnen mit Migrationshintergrund kumulieren. Es wäre wichtig, eine ähnliche Studie in Deutschland durchzuführen (als Beispiel für eine frühere Studie: vgl. Farrokhzad 2008). Auch bereits zu früheren Zeitpunkten, beispielsweise im Hinblick auf die Betreuung auf verschiedenen Qualifikationsstufen, sollten Studium und Lehre auf Barrieren und Diskriminierungsgefahren hin evaluiert und geeignete Gegenmaßnahmen entwickelt werden.

Fazit

Abschließend lässt sich feststellen, dass es vielfältige und auch gut ausgereifte Lösungsansätze für die Evaluation von Studium und Lehre gibt. Es existieren beispielsweise differenzierte Instrumente zur Erfassung der Studienmotivation, des Workloads sowie der Qualität von Lehrveranstaltungen. Da die genannten Verfahren bislang überwiegend auf subjektiven Einschätzun-

gen der Studierenden beruhen, liegt die Zukunft der Evaluationsforschung im Hochschulbereich in der Entwicklung objektiverer Maße, beispielsweise im Bereich der Kompetenzmessung. Der Sammelband informiert auch über zahlreiche Rahmenbedingungen der Studienaufnahme, Fachwahl sowie des Studienerfolgs bzw. des Studienabbruchs. Damit die vorgestellten Lösungsansätze nicht bloß isolierte Puzzlestücke bleiben, besteht eine Herausforderung für Akteure im Hochschulbereich darin, diese Ansätze in einer praktikablen Art und Weise zu verknüpfen. Für die Forschung wäre es beispielsweise interessant zu untersuchen, wie sich der Workload auf die Studienmotivation (bzw. auch umgekehrt), auf die Kompetenzen der Studierenden und auf die Studienabbruchquote auswirkt.

Der Sammelband erfüllt sein Ziel, Probleme und mögliche Lösungsansätze aufzuzeigen. Aus der Sicht der Rezensentin werden alle genannten Zielgruppen, d.h. Forschende und Studierende in den Bildungs- und Sozialwissenschaften, Evaluationsbeauftragte, Tätige in der Hochschulforschung sowie Hochschulleitungen und -management, durch die Beiträge angesprochen.

Literatur

- Farrokhzad, Schahrzad (2008): Erfahrungen, Strategien und Potenziale von Akademikerinnen mit Migrationshintergrund. In Hentges, Gudrun/Hinnenkamp, Volker/Zwengel, Almut (Hg.): Migrations- und Integrationsforschung in der Diskussion. Biografie, Sprache und Bildung als zentrale Bezugspunkte. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 303-322.
- Rindermann, Heiner/Baumeister, Antonia E. (2015): Validating the Interpretations of PISA and TIMSS Tasks: A Rating Study. In: International Journal of Testing, 15 (1), S. 1-22. doi:10.1080/15305058.2014.966911