

Christian Herfter & Robert Aust

Studentische Qualitätsurteile zu Ringvorlesungen. Überlegungen zu Theorie, Modell und Indikatoren.

Zusammenfassung

Evaluationen von Lehrveranstaltungen an Hochschulen sind seit den 1970er Jahren in Deutschland Gegenstand wissenschaftlicher Untersuchungen, wobei der Einbezug studentischer Qualitätsurteile im Evaluationsprozess kontroverse Diskussionen erfuhr (Aleamoni, 1994; Gold, 1996; Kromrey, 1994; Marsh, 2007). Insbesondere Heiner Rindermann (1997a, 1997b) konnte deren Bedeutsamkeit für die Hochschulforschung herausarbeiten. Dessen multidimensionales Modell der Lehrqualität (Rindermann, 1999) aufgreifend, verfolgen wir mit diesem Aufsatz zwei Ziele: Erstens soll als Beitrag zur Hochschul- und Evaluationsforschung hinterfragt werden, wie dieses Modell entlang eines bisher unberücksichtigten und dennoch bedeutsamen Veranstaltungstyps mit wechselnden Dozierenden, sog. Ringvorlesungen, genutzt werden kann. Darauf aufbauend stellen wir zweitens das Leipziger Inventar zur Evaluation von Ringvorlesungen (LIER) vor und bewerten dessen Generalisierbarkeit und Validität. Obgleich auf seine Vorläufigkeit und Exemplarität hinzuweisen ist, erhoffen wir uns eine produktive Resonanz im Feld der Lehrevaluation im Hinblick auf die Anwendung, (statistische) Überprüfung und Weiterentwicklung des Inventars.

Schlüsselwörter

Qualität universitärer Lehre, Lehrevaluation, Ringvorlesung, Fragebogen

1 Hinführung

Eine wesentliche Aufgabe der Hochschulen kann in der Ausbildung qualifizierter Nachwuchskräfte für Wissenschaft und Gesellschaft gesehen werden (vgl. Kopetz 2002, S. 94). Daraus folgt unmittelbar das Gebot „die Universitätslehre so gut wie möglich zu gestalten“ (Kopetz, 2002, S. 106). Eine möglichst optimale Gestaltung erfordert neben der inhaltlichen Realisierung der Lehre (1) auch ein Instrumentarium der Bewertung der erreichten Qualität (2). Insofern ist es wenig verwunderlich, dass beide Facetten als Aufgaben der Hochschulen im Hochschulrahmengesetz verankert sind (HRG §6).

Obwohl die Evaluation universitärer Lehre seit Mitte der 1980er Jahre fester Bestandteil der deutschen Hochschullandschaft ist, werden erst seit Mitte der 1990er Jahre verstärkt Evaluationsverfahren erprobt, welche die Studierenden als Urteilende der Lehre in den Blick nehmen (vgl. Rindermann, 1997a/2003). Viele etablierte Evaluationsmodelle und -verfahren fokussieren dabei entweder auf einzelne Lehrveranstaltungen, d.h. Seminar- und Vorlesungseinheiten (vgl. u.a. Brock & Keitel, 2016; Rindermann, 2003, S. 235), oder erheben Daten zu ganzen Studiengängen oder -modulen (vgl. z.B. die Beiträge in Großmann & Wolbring, 2016; Kuckartz & Rädiker, 2012).

Legt man – wie wir – eher Wert auf die Beschreibung einer längerfristigen Beziehung zwischen den Studierenden und der Hochschule, ermöglichen die Arbeiten Rindermanns (u.a. 1999) und sein „Modell der Lehrveranstaltungsqualität“ eine erste theoretische Annäherung. Dieses erlaubt mit seinen vier Variablen *Studierende*, *Dozierende*¹, *Rahmenbedingungen* und *Lehrerfolg* sowie damit verbundenen Wechselwirkungen eine Analyse der Qualität universitärer Lehrveranstaltungen und assoziierter Infrastruktur (z.B. technische Ausstattung, Platzangebot, Akustik usw.; Lenecke, 2005, S. 65). Empirisch erprobt wurde das Konzept Rindermanns an den Veranstaltungsformaten Seminar und Vorlesung, welche den Großteil universitärer Lehre ausmachen (Rindermann, 2003, S. 235).

Ausgangspunkt für unsere Überlegungen ist dagegen ein Veranstaltungsformat, das sich durch variierende Lehrpersonen und nur einen losen thematischen Bezug zu den „Methoden und Inhalten der Schul- und Unterrichtsforschung“ auszeichnet: die Ringvorlesung. Für dieses gängige, jedoch relativ seltene Format existieren weder national noch international eine systematische Analyse seiner Beschaffenheit und in der Folge auch keine Inventare zu seiner Evaluation aus Perspektive der Studierenden (vgl. zu dieser Leerstelle z.B. Schneider & Mustafić, 2015). Einzig der Beitrag von Eberhardt (2010) diskutiert dieses Konzept: Ihrer Beschreibung nach handelt es sich bei Ringvorlesungen um „klassische Veranstaltungsformen“, sie bleibt einer empirischen Untermauerung dieses Arguments jedoch ebenso schuldig wie Aussagen zur hohen Akzeptanz durch Dozierende und Lernende. Eine theoretische Erweiterung, welche das Konzept wechselnder Dozierender zu einem Thema einschließt und empirisch bspw. hinsichtlich Lehrerfolg oder Lehrveranstaltungsqualität evaluiert, erfolgte jedoch nicht.²

¹ Wir lehnen uns stark an die von Rindermann verwendeten Bezeichnungen an, schlagen jedoch vor, gender-sensible Begriffe konsequent zu verwenden, sodass wir beispielsweise statt *Dozent* das gender-neutrale Äquivalent *Dozierende* verwenden.

² Zwar benennt Eberhardt (2010, S. 273) wechselnde Dozierende (auch unterschiedlicher Fachgebiete) als Eigenschaften von Ringvorlesungen. Jedoch bezieht sie dies zum einen auf einen sehr

Mit diesem Beitrag verfolgen wir zwei Ziele und adressieren damit auch zwei Gruppen von Lesenden: *Erstens*, als Beitrag für die Hochschul- und Evaluationsforschung charakterisieren wir die Ringvorlesung als spezifischen Veranstaltungstyp und beschreiben ihre Merkmale im Kontext eines übergreifenden Qualitätsmodells. Hierbei wird auch auf die (kontrovers diskutierte) Eignung der Studierenden als Urteilende einzugehen sein. *Zweitens*, als Beitrag zur Lehr- und Evaluationspraxis erfolgt eine Modifizierung und Erweiterung des bewährten Heidelberger Inventars zur Lehrveranstaltungs-Evaluation (HILVE; Rindermann & Amelang, 1994) für die Konzipierung eines Instruments zur Evaluation dieses Veranstaltungstyps. Darauf aufbauend stellen wir das Leipziger Inventar zur Evaluation von Ringvorlesungen (LIER) vor und diskutieren es abschließend in Bezug auf seine Eignung.

2 Die Qualität von Ringvorlesungen

Ringvorlesungen zeichnen sich durch ein spezifisches Verhältnis von (In-)Varianz in Bezug auf Beteiligte, Inhalte und Rahmenbedingungen aus. Dies macht es erforderlich, ein adäquates und probates Instrument zu entwickeln, welches a) eine Bewertung der spezifischen Lehrveranstaltungsform erlaubt, dabei also u.a. die wechselnden Lehrenden berücksichtigt, und es b) ermöglicht, individuelle Rückmeldungen an einzelne Dozierende zu geben und somit deren fachliche und didaktischen Kompetenzen zu stärken und empirisch zu begleiten. Da es, wie bereits ausgeführt, noch keine fundierten theoretischen Rahmungen des Lehrveranstaltungskonzepts *Ringvorlesung(en)* gibt, scheint es naheliegend, sich über eine Kontrastierung zu den Merkmalen klassischer Vorlesungen dem Phänomen zu nähern.

Als *Vorlesung* wird eine Sequenz wissenschaftlicher Vorträge *einer* Person bezeichnet, die die *venia legendi* besitzt. In dieser Sequenz wird ein Sachverhalt abschnittsweise vermittelt und eine oder mehrere Fragestellung(en) werden zunehmend entfaltet. Den Ursprung hat diese Vermittlungsform in der antiken Rhetorik als erklärende Darstellung eines Textes. Trotz der mit dem Buchdruck entfallenen Notwendigkeit, Bücher vorzulesen, blieb und bleibt diese Form zur Vermittlung kanonischen Wissens fester Bestandteil der Hochschullehre. Den Studierenden kommt dabei die Aufgabe zu, den gegebenen Überblick zum Thema, aktuelle Forschungsstände und -perspektiven wie auch exemplarische Alltagsbezüge zu erkennen, aufzunehmen und darauf aufbauend eigene weiterführende Gedanken zu entwickeln (für den gesamten Abschnitt Apel, 1999, S. 63).

Vorlesungen werden als rhetorische Situation sehr stark an die Person und die Glaubwürdigkeit der oder des Vortragenden gebunden (Apel, 1999, S. 75). Bei *Ringvorlesungen*, d.h. einer Sequenz wissenschaftlicher Vorträge *mehrerer* Personen, muss die Zusammengehörigkeit der einzelnen Vorträge in jeweils größerem Maße hergestellt und aktualisiert werden, als dies für klassische Vorlesungen gelten kann. Dadurch vervielfältigt sich die Situation der Artikulation, d.h. der Struktur aus Einleitung, Darstellung, Argumentation und

engen thematischen Rahmen und es unterbleibt zum anderen eine empirische Fundierung ihrer Aussagen zum „Mehrwert“ dieser Veranstaltungsform. Der Beitrag hat daher in erster Linie die Charakteristik eines Erfahrungsberichtes.

Zusammenfassung, die in jeder einzelnen Veranstaltung durch die jeweils vortragende Person neu vollzogen werden muss. Damit kann – je nach der Offenheit der thematischen Rahmung der Ringvorlesung – der oben angesprochene Sequenzcharakter unterbrochen werden. Auf Ebene der Rhetorik ist dieser Charakter aufgrund der sehr individuellen ‚Vortragskunst‘ der Dozierenden (Apel, 1999, S. 72) ohnehin prekär. Bei einer Evaluation einer Ringvorlesung muss konzeptionell und methodisch daher auf allen Ebenen (Strukturen, Prozesse, Ergebnisse) berücksichtigt werden, dass jede Einzelveranstaltung für sich und gleichrangig deren Bedeutung für die Gesamtveranstaltung beurteilt werden können muss.³

Was spricht für die Bedeutsamkeit einer solchen hochschuldidaktisch herausfordernden, Veranstaltungsform? Die Auswahl *mehrerer* Dozierender ergibt sich aus deren fachlicher Expertise bezogen auf die Komplexität und interdisziplinäre Verwobenheit des thematischen Rahmens (so auch Eberhardt, 2010, S. 273). Viele Phänomenbereiche der Wissenschaft sind a) intern in hohem Maße funktional, theoretisch und methodisch ausdifferenziert oder/und weisen b) große interrelationale und auch interdisziplinäre Bezüge aus (vgl. bereits Stichweh, 1979), so dass übergreifende Expertisen nicht oder nur mit Einschränkungen von einer einzelnen Person zu erreichen sind. Ebenso ist die Tatsache, dass Wissenschaft und Forschung und somit auch deren Vermittlung international vernetzt und kollaborativ erfolgt (u.a. Hüther & Krücken, 2016), sodass es lohnenswert erscheint, den Charakter lokaler Bedingtheiten, des kollegialen Austausches und der fachlichen Diskussion perspektivierter Forschungsergebnisse im Format der Lehrveranstaltung zu erhalten. Hierfür erscheint die Ringvorlesung eine Reihe von Möglichkeiten bereit zu halten (Eberhardt, 2010, S. 275).

3 Qualität universitärer Lehre – die Perspektive der Studierenden

Versteht man das Studium als Prozess, so lassen sich Lehrveranstaltungen als die (quantitativ) dominierende, wenn auch nicht einzige strukturelle Komponente von Interaktionen zwischen Studierenden und der Hochschule (hier vor allem: Lehrenden) beschreiben (vgl. Herfter, 2014, S. 117). Hierbei wird mit Spady (1970, S. 77) von einem interaktionistischen Modell ausgegangen, womit eine situationsübergreifende, eher auf Dauer gestellte Beziehung zwischen den Studierenden (auch untereinander) und der Hochschule im Mittelpunkt steht, die durch (*zukunftsorientiertes*) *Vertrauen* und (*vergangenheitsorientierte*) *Vertrautheit* konstituiert wird (vgl. Herfter, 2014, S. 193). Vertrautheit kann in diesem Kontext als Retrospektion der gemachten Erfahrungen im universitären Kontext verstanden werden,

³ Unbenommen der situativen Realisierung konkreter Veranstaltungen geht es uns an dieser Stelle vielmehr darum aufzuzeigen, dass bezüglich der thematischen und artikulatorischen Offenheit einer Ringvorlesung *grundsätzlich* mehr Möglichkeiten und Notwendigkeiten der Gestaltung bestehen. Somit stellen sich, so unsere Annahme, erhöhte Anforderungen hinsichtlich der An- und Verknüpfungen von Einzelveranstaltungen und -themen an die Dozierenden und Studierenden.

die im Abgleich mit den Zielen, Erwartungen und Motiven der Studierenden zu Qualitätsurteilen führen (Austin & Vancouver, 1996, S. 338). Der so umrissene *Soll-Ist-Vergleich zwischen Erwartungen und Erfahrungen* lässt sich mit Caplan (1987, S. 249) hinsichtlich seiner Dimensionierung präziser fassen: Einerseits werden die erforderlichen Kompetenzen (Soll) mit den gestellten Anforderungen des Studiums (Ist) und andererseits individuelle Bedürfnisse (Soll) mit den vorhandenen Angeboten (Ist) verglichen. Diese Bewertungsprozesse münden in sog. Faktoren des subjektiven Studienerfolgs (vgl. Herfter, 2014, S. 194)⁴.

Auf der Folie dieses interaktionistischen Ansatzes lässt sich die Qualität der Lehre demnach als Facette des subjektiven Studienerfolgs – bezogen auf die studentischen Urteilsprozesse im Kontext einer einzelnen Lehrveranstaltungsreihe verstehen.

Bevor wir im folgenden Kapitel das multidimensionale Modell der Lehrqualität bzw. des Lehrerfolgs (Rindermann, 1999/2003/2009/2016)⁵ entlang struktureller Merkmale, prozessbezogener Qualitätsaspekte und deren Bewertung (vgl. Donabedian, 1980) vorstellen, gehen wir der Frage nach, inwiefern die Studierenden verlässliche und relevante Aussagen über die Qualität universitärer Lehre treffen können.

So merken Kritiker*innen in Bezug auf Studierende als Urteilende universitärer Lehre an, dass „studentische[n] Beurteiler einer Lehrveranstaltung weder im Ausmaß ihrer Zustimmung/Ablehnung noch in der Strukturiertheit ihrer Urteile übereinstimmen“ (Gold, 1996, S. 148) könnten. Auch wird aufgeführt, dass die studentische Kritik auf ein Instrument der Rückmeldung abzielte, nicht des Vergleiches „zwischen verschiedenen Veranstaltungen oder zwischen verschiedenen Lehrenden“ (Kromrey, 1994).

Aleamoni (1999), Marsh (2007) und Rindermann (1997a) zeigen auf, dass Studierende zu Unrecht als unqualifizierte Evaluator*innen von Hochschullehre gelten. Es zeigt sich, dass relevante Ergebnisse internationaler Forschung zu *student evaluation of teaching effectiveness* oder *students' evaluations of teaching* im deutschen Sprachraum nur unzureichend rezipiert worden sind und dass verschiedene (Meta-)Studien belegen, dass (gemittelte) studentische Urteile als objektiv, reliabel und valide gelten können.

Für die Beurteilung der Qualität von Ringvorlesungen ist es – wie eingangs ausgeführt – insbesondere bedeutsam, die verschiedenen Veranstaltungen kontinuierlich zu besuchen, um gleichermaßen individuelle wie auch vergleichende Rückmeldungen und Einschätzungen abgeben zu können. Auch hier haben Studierende im Vergleich zu möglichen externen Expert*innen einen erheblichen Vorteil: durch den Besuch vieler Veranstaltungen

⁴ Der Abgleich erforderlicher Kompetenzen entlang fachlicher Anforderungen des Studiums muss unseres Erachtens im Prüfungssystem verankert werden und kann erst darüber Bestandteil der Evaluation sein – Selbstauskünfte zum eigenen empfundenen Kompetenzniveau sind aus verschiedenen Gründen abzulehnen (vgl. Herfter, 2014, S. 26). In diesem Aufsatz konzentrieren wir uns daher auf studentische Bedürfnisse in Bezug auf Lehrangebote.

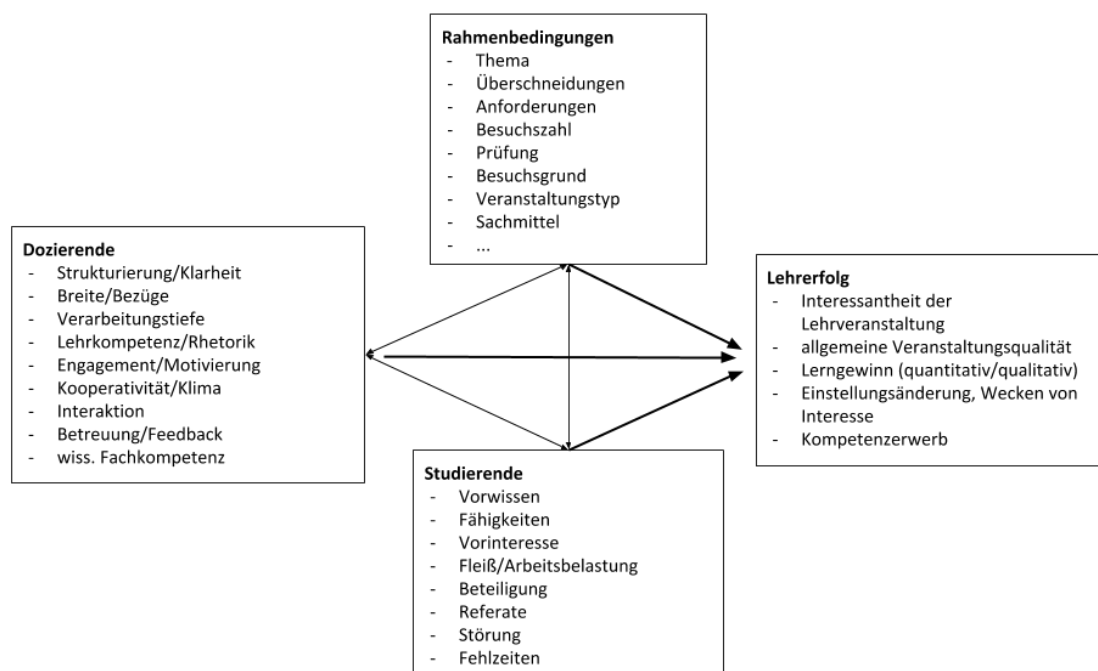
⁵ Rindermann bezeichnet sein Modell je nach zeitlicher Einordnung entweder u.a. als „(Münchner) multifaktorielles Modell der Lehrveranstaltungsqualität“ (1999, S. 363), „Multidimensionales Bedingungsmodell des Lehrerfolgs“ (2003, S. 236), „Multidimensionales Modell der Lehrqualität“ (2009, S. 66) oder „Bedingungs-Effekt-Modell der Lehrqualität“ (2016, S. 230). Unabhängig von den verschiedenen Bezeichnung beinhalten alle dargelegten Modelle die gleiche Trias um *Dozent* (2009, S. 66 *Lehrkraft*), *Studierende* und *Rahmenbedingungen* verknüpft mit der Produktvariable *Lehrerfolg*.

haben sie Erfahrungen mit guter und schlechter Lehre, können zwischen Lehrenden in Abhängigkeit vom Thema differenzieren. Auch sind sie die eigentlichen Adressat*innen der Lehre und müssen in der Folge auch im Hinblick auf die Akzeptanz der Evaluationsergebnisse einbezogen werden (Rindermann 2001, S. 63). Dem ließe sich als weiteres Argument hinzufügen, dass Lehr-Lern-Situationen als ko-konstruktive Prozesse zu verstehen sind, in denen Lernende und Lehrende gleichermaßen Verantwortung für die sozialen Aushandlungsprozesse und das (Re-)Konstruieren, Darstellen und Verteidigen des eigenen Wissens tragen (vgl. Chen, 2003, S. 20f.). Damit sind spezifische Erfahrungen verbunden, die durch professionelle Fremdbeurteilung unberücksichtigt bleiben würden. So ist es wenig verwunderlich, dass (selbst) im deutschen Hochschulraum die Urteile Studierender zunehmend zur Evaluation der Lehrqualität stärker berücksichtigt werden (Kuckartz, Dresing, Rädiker & Stefer, 2008, S. 11).

4 Das Multidimensionale Modell der Lehrqualität

Eines der im deutschsprachigen Raum etabliertesten Modelle zur Evaluation von Lehrveranstaltungsreihen, das den Interaktions- und Prozesscharakter des Studiums aufgreift und operationalisiert, ist das bereits benannte „Modell der Lehrveranstaltungsqualität“ Rindermanns (u.a. 1999). Zwar zielt sein Modell mit der Produktvariable Lehrerfolg auf eine „Leistungsbewertung“ der Dozierenden durch Studierende ab, jedoch rückt er explizit die Perspektiven der Studierenden als Adressat*innen auch als Mitgestaltende von Hochschullehre in den Mittelpunkt seines Modells (vgl. Rindermann, 2003). Wir stellen sein Modell zunächst vor und werden es im folgenden Kapitel entlang der o.g. Spezifika von Ringvorlesungen anpassen bzw. ergänzen. Erst durch diese Modifikationen von Modell und Erhebungsinstrument wird es möglich jene Aspekte zu evaluieren, die bei der Konzeption entlang ‚konventioneller‘ Vorlesungen und Seminare nicht berücksichtigt wurden (vgl. zur Anpassung von Evaluationsinstrumenten in der Hochschuldidaktik: Pfeiffer et al., 2015, S. 155).

Abbildung 1. Modell der Lehrveranstaltungsqualität nach Rindermann (1999, S. 363; eigene Darstellung; vgl. nahezu unverändert auch Rindermann, 2016, S. 230).



Die Wechselbeziehungen der drei Merkmalsbereiche Dozierende, Studierende und Rahmenbedingungen sind nach Rindermann (2009, S. 67) als interdependent zu beschreiben. Diese ‚Kausalbeziehungen‘ (Rindermann, 1999, S. 363f.) vollziehen sich entlang innerer und äußerer Verschränkungen. Je nach Lehrveranstaltung sind beispielsweise andere Modalitäten für Prüfungsleistungen oder Workload für Studierende vorgesehen, womit sich gleichzeitig auch unterschiedliche Verantwortungen und Abläufe der Wissensvermittlung und -diskussion durch Dozierende und Studierende ergeben. Vorlesungen beispielsweise zeichnen sich meistens durch eine größere Anzahl an teilnehmenden Studierenden aus, jedoch ist der Grad gemeinsamer sozialer Interaktion zwischen Lehrenden und Studierenden geringer als in Seminaren oder Tutorien. Lehrqualität ist somit als ein multidimensionales Phänomen charakterisierbar, welches sich durch vielseitige Wechselbeziehungen unterschiedlichster Variablen auszeichnet. Diese müssen in entsprechenden Evaluationsverfahren Berücksichtigung finden und können nicht losgelöst voneinander betrachtet werden (Rindermann, 2016, S. 229f.). Innerhalb dieser Beziehungen zwischen Dozierenden, Studierenden und Rahmenbedingungen vollzieht sich unter anderem jener einleitend skizzierte Prozess des Studierens als ein Etablieren von wechselseitigen Beziehungen zwischen Studierenden und Dozierenden in situationsspezifischen Settings. Mit dieser theoretischen Konstellierung ist es nun sowohl möglich Operationalisierungen vorzunehmen und jenen Prozess des Studierens empirisch zu erschließen, als auch Studierende als relevante Akteure im Prozess wie auch bei der Beurteilung von Lehrqualität und Lehrerfolg in den Mittelpunkt wissenschaftlichen Interesses zu rücken.

5 Das Leipziger Inventar zur Evaluation von Ringvorlesungen

Basierend auf a) den grundsätzlichen theoretischen Überlegungen zur Qualität universitärer Lehre, b) den konkreten Facetten und Zusammenhängen der Lehrveranstaltungsqualität sowie c) den Skalen und Items des *Heidelberger Inventar[s] zur Lehrveranstaltungs-Evaluation* (HILVE; vgl. Rindermann & Amelang, 1994) wurde ein Fragebogen für Studierende entwickelt (Aust & Herfter 2019)⁶, der im Zeitraum von drei Semestern im Rahmen einer Ringvorlesung der Universität Leipzig mit dem Titel „Methoden und Inhalte der Schul- und Unterrichtsforschung“ zum Einsatz kam. Die inhaltliche Ausrichtung sowie die Lehrenden blieben über diesen Zeitraum (nahezu) unverändert. Die behandelten Inhalte der Ringvorlesung sind auch in Form eines Sammelbandes publiziert worden, welcher die aufbauende und aufeinander verweisende Struktur widerspiegelt (Drinck, 2013). Unser Vorgehen ist dabei in erster Linie als *synkretistisch* und *theoretisch* zu bezeichnen, d.h. fußend auf „existierenden [...] Theorien und Lehrzielkonzeptionen“ sowie „der Auswahl und Kombination geeigneter Items und Skalen“ aus „vorhandene[n] Verfahren und Ergebnisse der Lehrvaluationsforschung“ (Rindermann, 1999, S. 358). Dabei erwies sich die bereits oben erörterte Leerstelle zu Evaluationskonzepten von Ringvorlesungen als besondere inhaltliche und methodische Herausforderung. Wir begegneten dieser Herausforderung durch wiederholte Diskussionen im Kreis der Lehrenden zu den relevanten Kriterien guter Lehre im Kontext dieses spezifischen Veranstaltungsformats. Dabei ging es darum, bedeutsame Items auszuwählen bzw. Ergänzungen dort vorzunehmen, wo die diskursiv erarbeiteten Kriterien und Items über die Fachliteratur hinausgehen. Nachfolgend wird es darum gehen, die Generalisierbarkeit der veranstaltungsspezifisch entwickelten Kriterien und Items zu plausibilisieren und – soweit möglich – theoretisch zu begründen. Insgesamt kann der Anspruch jedoch nur sein, das Leipziger Inventar vorzustellen und so eine wissenschaftliche Auseinandersetzung zur Evaluation von Ringvorlesungen anzuregen.

5.1 Entwicklung des Fragebogens

In der nachfolgenden Tabelle stellen wir zunächst alle in Bezug auf HILVE übernommenen⁷, modifizierten und ergänzten Items als Übersicht dar, bevor wir diese Arbeitsschritte jeweils einzeln ausführen und begründen. Aufgeführt sind dabei weiterhin die Zuordnung der Items des LIER zu den Dimensionen aus dem Modell der Lehrqualität.

⁶ Der Fragebogen steht vollumfänglich online unter <https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa2-357101> zur Verfügung und kann für die eigene Evaluation(sforschung) verwendet werden.

⁷ Neben der Formulierung der Fragen wurden auch die Antwortformate angepasst. Statt mehrstufiger, abgestufter Skalen kommen visuelle Analogskalen zum Einsatz. Für eine Darstellung der Vorzüge eines solchen Vorgehens vgl. z.B. Greinöcker (2009).

Tabelle 1. Übersicht der Items aus dem Leipziger Inventar zur Evaluation von Ringvorlesungen. Zeilenweise ergibt sich eine Zuordnung zu den zentralen Variablen des Modells der Lehrqualität, spaltenweise eine Aussage zur ‚Herkunft‘ der Fragestellungen in Bezug auf HILVE.

	übernommene Items	modifizierte Items	ergänzte Items
Dozierende	waren gut vorbereitet	erklärten Inhalte verständlich	weckten Interesse
	Anregung zu kritischer Auseinandersetzung	Vermittlungstempo	beantworteten Fragen zufriedenstellend
	förderten Fragen und Beiträge		Vorlesungen begannen pünktlich
	gaben Raum für Diskussionen		Vorlesungen endeten pünktlich
Rahmenbedingungen	Medien hilfreich für Verständnis	Kenntnisse über die Ziele der Ringvorlesung	Verständnis Ziele der Ringvorlesung
	Raum für die Veranstaltung geeignet	inhaltliche Gesamtkonzept der RVL verständlich	Bezüge einzelner Vorlesungen erkennbar
	Akustik	Folien hilfreich für Verständnis	angenehme Arbeitsatmosphäre
		ergänzende Materialien hilfreich	Anzahl der Themen, Umfang der Inhalte
Studierende	Interesse am Thema	Fleiß/Arbeitshaltung	Aufarbeitung der Inhalte
			Wunsch nach Unterstützung zur Aufbereitung
Lehrerfolg	Zufriedenheit gesamt	fünf Vorlesungen für positiven Eindruck	Relevanz der Themen
		fünf Vorlesungen für negativen Eindruck	Wünsche für die Ringvorlesung

Für *Dozierende* als Merkmalsbereich, geordnet nach dem Modell von Rindermann, kamen für die Erarbeitung des LIER, mit Blick auf die didaktischen Kompetenzen der Lehrenden sowie die Anforderungen an die Artikulationssituation, die Items aus den Bereichen (1) *Strukturierung / Klarheit*, (2) *Breite / Bezüge*, (3) *Verarbeitungstiefe* und (4) *Lehrkompetenz / Rhetorik* in Frage. Mit der ersten Dimension wird der nachvollziehbare Aufbau einer Veranstaltung beschrieben, welcher Verständnisprozesse und die kognitive Strukturierung von zu erlernenden Inhalten und deren späteren Abruf umfasst und nach Rindermann (1998, S. 201f.) die wichtigste Bezugsdimension zur Beurteilung von Lehre darstellt. Mit der zweiten Dimension werden Beispiele und Praxisbezüge bezeichnet, welche die Strukturierung der Lernprozesse erleichtern und in der Folge Lernerfolge ermöglichen (Rindermann, 1998, S. 201f.). Mit der dritten von uns übernommenen Dimension wird betont, dass sich die Anregung der aktiven Anwendung und Bewertung kanonischer Wissensbestände durch die Studierenden positiv auf den Lernerfolg auswirkt (Rindermann, 1999, S. 365). Unter der vierten und letzten Dimension wird „eine allgemeine Fähigkeit zu lehren“ subsumiert –

solch eine didaktische Kompetenz äußert sich in guter Vorbereitung, Sprachfähigkeit und Fähigkeiten der Komplexitätsreduktion (Rindermann, 1999, S. 365). Ergänzt haben wir die von Rindermann konzipierten Items um solche, welche die aufgezeigten interaktionistischen Merkmale universitärer Lehre stärker bzw. vertiefter berücksichtigen: So formulierten wir Fragen zu Interessensgenerierung (lässt sich der Dimension *Engagement / Motivierung* zuordnen; Rindermann, 1999, S. 366) und Zufriedenheit in Diskussionen (Dimension *Interaktion*; Rindermann, 1999, S. 366). Ebenso ergänzten wir Angaben zur Pünktlichkeit der Dozierenden in Ringvorlesungen, eine Dimension, die aus unserer Sicht im Zusammenhang mit eng getakteten Studententafeln und Wegezeiten nicht nur ein Ausdruck der Berücksichtigung wechselseitiger Erwartungen, sondern auch eine studienorganisatorische Notwendigkeit ist. Nicht berücksichtigt werden damit die Dimensionen *Kooperativität / Klima* – welche stärker die Lehrenden als die Lehrveranstaltung in den Blick nimmt –, *Betreuung / Feedback* – womit in erster Linie Rückmeldungen zu studentischen Referaten und Hausarbeiten bezeichnet sind, die im Format der Ringvorlesung keine Rolle spielen – und *wissenschaftliche Fachkompetenz* – die u.E. zwar zu den zentralen Voraussetzungen universitärer Lehre gehört, aber durch etablierte Peer-Verfahren und nicht (allein) durch Studierende bewertet werden kann (vgl. Herfter, 2014, S. 90f.).

Im Merkmalsbereich *Studierende* wird in erster Linie deren Handeln beschrieben, wobei dieses eng an kognitive Konzepte wie Vorbildung, Interesse und Engagement gekoppelt wird (vgl. Rindermann, 1999, S. 362). Wir wählten aus dem Modell von Rindermann die drei Dimensionen (1) *Fleiß / Arbeitsbelastung*, (2) *Beteiligung* und (3) *Vorinteresse* aus. Zur ersten Dimension zählte in erster Linie die Vor- sowie Nachbereitung von (besuchten) Lehrveranstaltungen. Sie stellt nach Rindermann (1999, S. 367f.) eine eigene Dimension dar, die individuelle Arbeitshaltungen und als Mittelwert die Arbeitsbelastung durch die Veranstaltung widerspiegelt. Mit Blick auf diese mittlere Arbeitsbelastung als ein Ausdruck der gestellten *Anforderungen* koppelten wir diese über zwei ergänzende Items der entsprechenden Prozessvariablen der Rahmenbedingungen (Rindermann, 1999, S. 369). Neu konzipiert wird die Items „Aufarbeitung der Inhalte“ (als Ausdruck individueller Arbeitshaltungen und mittlerer Arbeitsbelastung) und „Wunsch nach Unterstützung zur Aufbereitung“ (als Ausdruck einer möglichen Überforderung). Die ursprünglichen Items aus dem Merkmalsbereich *Anforderungen* modifizierten wir, um die aus unserer Sicht negativen Formulierungen der Fragestellungen zu ‚neutralisieren‘. Die an *Fleiß* anknüpfende und zweite von uns übernommene Dimension *Beteiligung* stellt das studentische Pendant zu *Verarbeitungstiefe* sowie *Interaktion* bei den Dozierenden dar und wird analog begründet (Rindermann, 1999, S. 368). Das *Vor-Interesse* am Thema, d.h. eine wertschätzende Zuwendung, ist eine wichtige Variable in Bezug auf Lernmotivation, Beteiligung und Fleiß damit aber auch eine sog. Bias-Variable in Bezug auf die Beurteilung der Lehrveranstaltungsqualität (Rindermann, 1999, S. 367); wir verankern es dennoch (nur) als studentisches Merkmal und nicht als Rahmenbedingung. Nicht oder nur teilweise berücksichtigt wurden die Dimensionen *Störungen* – teils erfasst durch Pünktlichkeit und Akustik –, *Fehlzeiten*, *Referate* – aufgrund des Veranstaltungsformates – sowie *Vorwissen* und *Fähigkeiten*, die nicht oder nur unzureichend in freiwilligen Evaluationsbögen erfasst werden können (vgl. den folgenden Abschnitt zum Lehrerfolg).

Mit Rindermann (1999, S. 362) beschreiben wir *Rahmenbedingungen* als Prozessvariable „günstige [...] in Form von ausreichender Zeit, Platz, adäquater Anforderungshöhe und inhaltsangemessener Veranstaltungsformen“. Hier wurden (1) *Sachmittel* (Medieneinsatz, Technik, Material, Skript etc.) sowie (2) *Anforderungen* übernommen sowie frei hinzufügbare Items des HILVE für Rahmenbedingungen (Materialien/Skript, Raum, Akustik – ergänzt um Arbeitsatmosphäre) ausgewählt. Den Aspekt der ‚Materialien‘ differenzierten wir hinsichtlich der (Präsentations-)Folien und ergänzender Materialien wie Begleittexte, Handouts etc. Die Dimension *Anforderungen* stellt – als Herausforderung studentischer Kompetenzen – ein zentrales Merkmal zur Beschreibung der Qualität von universitären Lehrveranstaltungen dar (Anspruch und Umfang des Inhalts, Tempo). Obgleich Strukturierung und Klarheit (als Dimension des Dozierendenverhaltens) eine zentrale Verantwortung für erfolgreiche Lehrveranstaltung zugesprochen werden, wird deren Bezug zur Transparenz und Verständlichkeit von Zielen im Modell von Rindermann nicht systematisch entfaltet. Insbesondere für Ringvorlesungen, deren Zusammengehörigkeit bzw. Integration durch eine gemeinsame Zielsetzung sich nicht auf eine Kontinuität der Vortragenden Person oder Themen stützen kann, gewinnt die Dimension *Ziele* besonderes Gewicht für eine Qualitätsbeurteilung (entlang des Abgleichs von Erwartungen und Erfahrungen). Entlang dieser Überlegungen modifizierten wir ein ‚frei hinzufügbares‘ Item des HILVE zur Transparenz bzw. Kenntnis der Ziele und ergänzten diesen Aspekt um die Verständlichkeit der Ziele sowie die Bezüge zwischen den einzelnen Veranstaltungen. Nicht berücksichtigt wurden im Fragebogen die Dimensionen thematischer Überschneidungen (positiv gewendet in ‚Ziele‘ abgebildet), Personalmittel, Beratungen, Gebühren, Besuchsgrund und Veranstaltungstyp, da diese Eigenschaften für die untersuchte Ringvorlesung invariant bzw. nicht vorhanden waren.

In Bezug auf den Lehrerfolg haben wir die Effektvariablen *Lerngewinn*, *Einstellungsänderungen* und *Kompetenzerwerb* zunächst ausgeschlossen, da hier die Freiwilligkeit der studentischen Beurteilung der Lehrveranstaltung und der verbindliche Nachweis intersubjektiv-feststellbarer Qualifikation sowohl in inhaltlich-methodischer (vgl. z.B. Weinert, 1999, S. 26) als auch in organisatorisch-struktureller Sicht (vgl. Reis & Ruschin, 2008, S. 55) auf ungünstige Weise verschränkt sind. Facetten der subjektiven Sinngebung des Studiums sind dagegen als *Relevanz* in den Items des Lehrerfolgs verankert. Wir beschränken uns in der Beschreibung des Lehrerfolgs von Ringvorlesungen durch studentische Evaluation in der Folge auf ein globales Urteil der Gesamtzufriedenheit⁸. Zur Einlösung des Anspruchs, jede Einzelveranstaltung für sich und gleichrangig deren Bedeutung für die Gesamtveranstaltung beurteilen zu können, ergänzten wir zwei Freitextfragen, in denen besondere positive wie negative Erfahrungen mit bis zu fünf bestimmten Vorlesungen formuliert werden können. Zur Erhöhung des partizipativen und mitgestaltenden Charakters der Evaluation ergänzten wir das offene Item „Wünsche für die Ringvorlesung“.

⁸ Anders als Rindermann (1999, S. 370) betrachten wir *Interessantheit* (als Wechselspiel von Dozierenden- und Studierendenhandeln) nicht als Dimension der Lehrqualität, sondern von Struktur- und Prozessmerkmalen.

5.2 Statistische Reflexion

Bedingt durch eine Freiwilligkeit der Teilnahme sowie nur indirekter Prüfungsrelevanz (über eine in den Seminaren entwickelte und betreute Hausarbeit mit forschend-empirischer Ausrichtung) konnte nur eine relativ kleine Gruppe von Studierenden (Sommersemester 2012 n=42, Wintersemester 2012/2013 n=35, Sommersemester 2013 n=31) als kontinuierlich Hörende gewonnen werden. Diese Studierenden wurden in der letzten Veranstaltung gebeten, ihre Qualitätseinschätzung auf Papier zu bringen; damit erreichten wir einen nahezu vollständigen Rücklauf.

Wohl wissend um die eingeschränkte Aussagekraft der erhobenen Daten dreier kleiner Kohorten einer einzelnen, inhaltlich spezifischen und örtlich gebundenen Veranstaltung, können wir nur eine vorsichtige Analyse zur Generalisierbarkeit der Evaluationsergebnisse und damit der Güte des Fragebogens vornehmen. Zunächst sei daran erinnert, dass nach aktuellem Kenntnisstand der generelle Zweifel an der Objektivität, Reliabilität und Validität von Studierendenrückmeldungen unbegründet ist (vgl. Kapitel 3). Es ist demnach also „nur“ zu zeigen, dass die veranstaltungsbezogenen Urteile, unter der Annahme unveränderter Themen, Gestaltung, Lehrmethoden usw., über die drei Kohorten hinweg hinsichtlich zentraler Tendenzen⁹ stabil sind. Denn dann würde das Instrument eine verlässliche Qualitätsbeschreibung in den o.g. Merkmalsbereichen erlauben.

Die untersuchten Items konnten durch den Einsatz von visuellen Analogskalen zwar metrisch skaliert werden, sind jedoch in etwa der Hälfte der Fälle nicht normalverteilt. Dies betrifft die Fragen zur Anzahl der Themen, zum Raum für Diskussionen, zur Aufbereitung der Vorlesungsinhalte, zur Nützlichkeit der Präsentationsfolien, zur Akustik, sowie zum pünktlichen Anfang und Ende der Veranstaltungen (Kolmogorov-Smirnov-Test, $p < .05$). Um eine einheitliche Analyse zu ermöglichen, entschieden wir uns daher für den Mann-Whitney-U-Test um zu prüfen, ob die zentralen Tendenzen der drei unabhängigen Stichproben verschieden sind. Die Voraussetzungen des Tests werden erfüllt, da die unabhängige Variable (Kohortenzugehörigkeit) dichotom ist und die abhängigen Variablen (Qualitätsrückmeldungen entlang der o.g. Items) aus unabhängigen Variablen stammen, mindestens ordinal skaliert sind und zudem eine ähnliche Form der Verteilung aufweisen. Letzteres wurde durch einen Abgleich der Boxplots und Histogramme geprüft (vgl. Fagerland & Sandvik, 2009). Die nachfolgende Tabelle zeigt die Testergebnisse aller paarweisen Vergleiche sowie Effektstärken und Mittelwerte.

⁹ Aus diesem Grund beschränken wir unsere Analysen auf mindestens ordinal skalierte Items; offene Antwortformate und Ja/Nein-Fragen lassen wir außen vor.

Tabelle 2. Ergebnisse des Kohortenvergleichs mittels Mann-Whitney-U-Test. Modifizierte Items werden mit einem Stern markiert, neu entstandene Items durch zwei Sterne. Da die Stichprobengröße N jeweils größer als 30 ist, werden die asymptotische Signifikanz (Sig.) und die Effektstärke (r, Korrelationskoeffizient nach Pearson) ausgewiesen. Signifikante Ergebnisse ($p < .05$) und deren Effektstärke sind fett hervorgehoben. Für die Interpretation der Ergebnisse geben wir zusätzlich noch die Mittelwerte an. Das Optimum ist bei fast allen Items 0 (Sehr/Immer), der schlechteste Wert 100 (Nicht/Nie). Ausnahme stellen die Items Anzahl der Themen, Umfang der Inhalte und Raum für Diskussionen dar: hier liegt das Optimum bei 50 – 0 ist ‚zu viel‘, 100 ‚zu wenig‘.

	SS12 - WS1213				SS12 - SS13				WS1213 - SS13				Mittelwerte		
	N	Z	Sig.	r	N	Z	Sig.	r	N	Z	Sig.	r			
Interesse am Thema	77	-0,84	0,401	0,096	73	-1,26	0,209	0,147	66	-2,13	0,033	0,262	47,7	51,5	42,0
Bezüge einzelner Vorlesungen	75	-0,64	0,522	0,074	72	-0,24	0,810	0,028	63	-0,61	0,545	0,077	37,9	35,4	39,4
Anzahl der Themen	74	-1,31	0,192	0,152	69	-0,92	0,358	0,111	63	-0,53	0,598	0,067	38,9	43,9	43,0
Umfang der Inhalte	75	-0,52	0,600	0,060	69	-0,80	0,425	0,096	62	-0,08	0,938	0,010	37,0	40,2	41,4
D. weckten Interesse**	77	-0,90	0,370	0,103	72	-2,77	0,006	0,326	65	-1,51	0,130	0,187	57,3	52,9	45,8
D. erklärten Inhalte verständlich*	76	-0,46	0,646	0,053	71	-0,03	0,972	0,004	65	-0,41	0,683	0,051	33,5	36,1	32,1
D. waren gut vorbereitet	75	-3,04	0,002	0,351	70	-1,52	0,128	0,182	63	-1,46	0,145	0,184	13,2	23,6	18,6
Anregung zu kritischer Auseinandersetzung	74	-0,86	0,391	0,100	68	-1,02	0,310	0,124	62	-0,15	0,882	0,019	50,3	45,4	44,4
D. förderten Fragen und Beiträge	76	-1,88	0,060	0,216	70	-1,63	0,103	0,195	62	-0,29	0,772	0,037	27,2	36,4	33,9
D. beantworteten Fragen zufriedenstellend**	75	-1,92	0,055	0,222	69	-2,63	0,008	0,317	62	-0,91	0,361	0,116	17,6	25,9	31,8
D. gaben Raum für Diskussionen	75	-1,72	0,086	0,199	69	-2,28	0,023	0,274	60	-0,39	0,693	0,050	41,0	50,6	52,2
Vermittlungstempo*	76	-1,61	0,108	0,185	70	-1,97	0,049	0,235	64	-0,16	0,871	0,020	41,5	33,2	32,5
Aufbereitung der Inhalte: Gar nicht**	71	-1,76	0,079	0,209	64	-1,17	0,242	0,146	57	-0,43	0,664	0,057	55,4	40,3	46,5
A.d.I.: Vorlesungsfolien herunterladen**	77	-1,00	0,315	0,114	71	-1,28	0,200	0,152	64	-0,61	0,541	0,076	24,1	24,0	31,6
A.d.I.: Vorlesungsfolien lesen**	76	-0,59	0,556	0,068	70	-0,16	0,872	0,019	64	-0,47	0,637	0,059	47,7	42,1	46,0
A.d.I.: Literatur recherchieren**	76	-1,81	0,070	0,208	70	-2,74	0,006	0,327	64	-0,90	0,371	0,113	92,3	83,7	78,6
A.d.I.: Literatur lesen**	76	-1,97	0,048	0,226	70	-2,81	0,005	0,336	64	-0,70	0,485	0,088	91,7	82,2	79,6
A.d.I.: eigene Mitschriften bearbeiten**	75	-1,62	0,105	0,187	69	-2,16	0,031	0,260	64	-0,41	0,681	0,051	67,9	56,1	52,4
A.d.I.: Video-stream**	73	-1,56	0,118	0,183	69	-1,50	0,133	0,181	62	-0,01	0,988	0,001	84,4	74,5	73,6
Meiden hilfreich für Verständnis	76	-1,60	0,109	0,184	71	-0,42	0,675	0,050	65	-2,05	0,040	0,254	22,3	14,0	20,5

Folien hilfreich für Verständnis*	75	-0,16	0,877	0,018	69	-1,90	0,058	0,229	64	-1,97	0,048	0,246	17,2	13,8	19,8
ergänzende Materialien hilfreich*	63	-0,31	0,757	0,039	55	-0,97	0,330	0,131	54	-0,53	0,599	0,072	35,8	36,5	32,7
Akustik	75	-0,29	0,771	0,033	70	-0,19	0,852	0,023	65	-0,35	0,726	0,043	19,7	17,5	15,4
Vorlesungen begonnen pünktlich**	74	-1,41	0,158	0,164	70	-2,34	0,019	0,280	64	-1,11	0,265	0,139	8,6	9,2	13,6
Vorlesungen endeten pünktlich**	73	-2,07	0,038	0,242	69	-3,89	0,000	0,468	62	-2,22	0,026	0,282	7,3	13,5	26,7
angenehme Arbeitsatmosphäre**	73	-1,02	0,309	0,119	68	-1,51	0,131	0,183	63	-0,66	0,508	0,083	40,1	34,7	30,4
Zufriedenheit gesamt	75	-0,65	0,515	0,075	70	-3,02	0,002	0,361	61	-1,40	0,162	0,179	51,7	47,1	38,8

Das erste augenscheinliche Ergebnis ist, dass die meisten signifikanten Änderungen zwischen den zeitlich nicht aufeinanderfolgenden Kohorten SS12 und SS13 auftreten. Betrachtet man diese Differenzen inhaltlich (entlang der Veranstaltungsmittelwerte), so steigen das inhaltliche Interesse und die Vorbereitung der Studierenden, die Zufriedenheit mit dem Vermittlungstempo sowie der Gesamtzufriedenheit. In diesem Vergleich werden aber auch organisatorische Probleme thematisiert: die Beantwortung von Fragen wird im Zeitraum von drei Semestern kontinuierlich schlechter bewertet, gleiches gilt für den Raum für Diskussionen, sowie die Pünktlichkeit (Beginn und Ende). Es handelt sich dabei um schwache bis mittlere Effekte. Da keine weiteren Evaluationsdaten vorliegen kann nun nicht final erklärt werden, ob diese Tendenzen auf Veränderungen der Veranstaltungsqualität oder auf ein davon unabhängiges und verändertes Rückmeldeverhalten zurückzuführen sind. Aufgrund der Kontinuität der beschriebenen Veränderungen unabhängiger Studierendengruppen, und der relativ geringen Effektstärken spricht u.E. einiges dafür, von einer geringen, aber feststellbaren Veränderung der „realen“ Lehrqualität auszugehen, die im Instrument korrekt abgebildet wird.

Diskontinuierliche Beschreibungen der Veränderungen sind ebenfalls festzustellen: Dies betrifft das persönliche Interesse am Thema (am geringsten im WS12/13, am höchsten im SS13), die Vorbereitung der Dozierenden (ebenfalls am geringsten im WS12/13, am besten im SS12), sowie die Einschätzung zur Nützlichkeit der Folien und eingesetzten Medien (am geringsten im SS13, am höchsten im WS12/13). Wie bereits eben dargelegt ist es uns auf dieser Datengrundlage nicht möglich, diese Differenzen den variierenden Qualitäten oder vermeintlich unzuverlässigen Studierendenrückmeldungen zuzuordnen.

Um die Güte des Instruments abschließend beurteilen zu können, bedarf es demnach weiterer Erprobung: im Kontext anderer thematischer Bezüge zur Plausibilisierung der Generalisierbarkeit sowie – idealerweise – begleitet von anderen Evaluationsverfahren als externen Kriterien der Validität (vgl. Rindermann, 1997, S. 220).

6 Resümee und Ausblick

Das Leipziger Inventar zur Evaluierung von Ringvorlesungen steht am Ende eines analytischen Prozesses: Aus einer standortgebundenen Gütebewertung einer Vorlesung mit wechselnden Dozierenden zu Methoden und Inhalten der Schul- und Unterrichtsforschung

heraus entstand die Idee, diese Evaluation selbst forschend zu reflektieren und zu generalisieren. Dabei zeigte sich, dass das Veranstaltungsformat der Ringvorlesung innerhalb der Hochschul- und Evaluationsforschung hinsichtlich ihrer spezifischen (In-)Varianz in Bezug auf Beteiligte, Inhalte und Rahmenbedingungen noch kaum berücksichtigt wurde. Es galt daher zu prüfen, welche Theoretisierungen und Modellierungen eine Beschreibung der situationsübergreifenden und andauernden Beziehung zwischen den Studierenden und den Lehrenden ermöglichen und eine empirische bewährte Operationalisierung der Lehrqualität aus Studierendensicht zur Verfügung stellen. Mit dem „Modell der Lehrveranstaltungsqualität“ (Rindermann, 1999) und dem Heidelberger Inventar zur Lehrveranstaltungs-Evaluation (HILVE; Rindermann & Amelang, 1994) steht ein solcher Forschungszusammenhang bereits zur Verfügung, der aber die spezifischen Merkmale von Ringvorlesungen, insbesondere deren prekäre Kontinuität, nicht systematisch berücksichtigt. Die im Text vorgeschlagenen und plausibilisierten Erweiterungen und Modifikationen des Fragebogens für Studierende können folglich als (Reflexions-)Beitrag für die Hochschul- und Evaluationsforschung gelten.

Darüber hinaus stellen wir den entwickelten Fragebogen Praktiker*innen und Lehrvaluationsforscher*innen zur Verfügung, um diesen einzusetzen, kritisch zu prüfen und weiterzuentwickeln. Beim gegenwärtigen Entwicklungsstand des LIER muss auf dessen Vorläufigkeit und Exemplarität verwiesen werden: Obgleich wir die vorgenommenen Veränderungen am Fragebogen vom Qualitätsmodell theoretisch ableiten, auf Item-Ebene plausibilisieren und im Kreis der Lehrenden kommunikativ validierten, kann zunächst nicht davon ausgegangen werden, dass der veränderte Fragebogen weiterhin (im statistischen Sinne) valide und reliabel ist. Erste Analysen (Kap. 5.2) deuten jedoch auf eine mögliche Validität und Generalisierbarkeit des Inventars zu Erhebung der Lehrqualität von Ringvorlesungen hin.

Es wäre aus unserer Sicht wünschenswert, dass dieser Beitrag dazu anregt, Modelle der Lehrveranstaltungsevaluation auch an „ungewöhnlichen“ Formaten, wie Ringvorlesungen, aber auch Team Teaching, Tutorien, Exkursionen etc. zu hinterfragen. Ebenfalls wünschenswert wäre es, wenn das vorgestellte Inventar in thematisch variierenden Kontexten eingesetzt und umfassend statistisch geprüft würde, sowie eine Weiterentwicklung – z.B. hinsichtlich obligatorischer und frei hinzufügbare Items (analog zum HILVE) oder zur systematischen Berücksichtigung interdisziplinärer Bezüge – erführe. Gerade letzteres konnten wir nicht systematisch berücksichtigen, da in unserem Anwendungsfall die wissenschaftlichen Perspektiven und Untersuchungsgegenstände relativ homogen ausfielen.

Literatur

- Aleamoni, L. M. (1999). Student Rating Myths Versus Research Facts From 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13(2), 153-166.
- Apel, H. J. (1999). „Das Abenteuer auf dem Katheder“. Zur Vorlesung als rhetorische Lehrform. *Zeitschrift für Pädagogik*, 45(1), 61–79.
- Aust, R. & Herfter, C. (2019). Das Leipziger Inventar zur Evaluierung von Ringvorlesungen (LIER) – Fragebogen. Zugriff am 15.10.2019. Verfügbar unter <https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa2-357101>

- Austin, J. T. & Vancouver, J. B. (1996). Goal Constructs in Psychology: Structure, Process, and Content. *Psychological Bulletin*, 120 (3), 338–375.
- Braun, E., Gusy, B., Leidner, B. & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, 54 (1), 30–42.
- Brock, M. & Keitel, J. (2016). Responsive Evaluation universitärer Lehrveranstaltungen. In *Zeitschrift für Evaluation*, 15(2), 269–292.
- Caplan, R. D. (1987). Person-Environment Fit Theory and Organizations: Commensurate Dimensions, Time Perspectives, and Mechanisms. *Journal of Vocational Behavior*, 31, 248–267.
- Centrum für Hochschulentwicklung. (2009). CHE-Studierendenfragebogen. Zugriff am 03.07.2018. Verfügbar unter http://www.che.de/downloads/Studierenden_Fragebogen_2009.pdf.
- Chen, C. (2003). A Constructivist Approach to Teaching: Implications in Teaching Computer Networking. *Information Technology, Learning, and Performance Journal*, 21 (2), 17–27.
- Cramer, C. (2010). Kompetenz Lehramtsstudierender. Grenzen und Perspektiven selbst-eingeschätzter Kompetenzen in der Lehrerbildungsforschung. In A. Gehrman, U. Hericks & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 85–97). Bad Heilbrunn: Klinkhardt.
- Eberhardt, U. (2010). Ringvorlesungen als Diskussionsforum und Instrument hochschuldidaktischer Weiterbildung. In *Neue Impulse in der Hochschuldidaktik* (S. 273–277). Wiesbaden: Springer VS.
- Eccles, J. S. & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53 (1), 109–132.
- Fagerland, M. W. & Sandvik, L. (2009). The Wilcoxon–Mann–Whitney test under scrutiny. *Statistics in Medicine*, 28(10), 1487–1497
- Greinöcker, A. (2009). Der Einsatz visueller Analogskalen (VAS) in Online-Befragungen. In Weichbold, M., Bacher, J. & Wolf, C. (Hrsg.). *Umfrageforschung. Herausforderungen und Grenzen* (S. 131–155). Wiesbaden: Springer VS.
- Großmann, D. & Wolbring, T. (Hrsg.) (2016). *Evaluation von Studium und Lehre. Grundlagen, methodische Herausforderungen und Lösungsansätze*. Wiesbaden: Springer VS.
- Harring, M., Böhm-Kasper, O., Rohlf, C. & Palentien, C. (2010). Peers als Bildungs- und Sozialisationsinstanzen. Eine Einführung in die Thematik. In Harring, M., Böhm-Kasper, O., Rohlf, C. & Palentien, C. (Hrsg.). *Freundschaften, Cliques und Jugendkulturen. Peers als Bildungs- und Sozialisationsinstanzen*. Wiesbaden: Springer VS. S. 9–19.
- Herfter, C. (2014). *Qualität universitärer Bildung. Theoretische und empirische Explorations zur Perspektive der Studierenden*. Leipzig: Leipziger Universitätsverlag.
- Hüther, O. & Krücken, G. (2016). *Hochschulen. Fragestellungen, Ergebnisse und Perspektiven der sozialwissenschaftlichen Hochschulforschung*. Wiesbaden: Springer.
- Kopetz, H. (2002). *Forschung und Lehre. Die Idee der Universität bei Humboldt, Jaspers, Schelsky und Mittelstraß*. Wien: Böhlau.

- Kromrey, H. (1994). Evaluation der Lehre durch Umfrageforschung? In: Mohler, P. P. (Hrsg.). *Universität und Lehre. Ihre Evaluation als Herausforderung an die empirische Sozialforschung* (S. 91–113). Münster: Waxmann.
- Kuckartz, U., Dresing, T., Rädiker, S. & Stefer, C. (2008). *Qualitative Evaluation*. Wiesbaden: Springer VS.
- Kuckartz, U. & Rädiker, S. (Hrsg.) (2012). *Erziehungswissenschaftliche Evaluationspraxis*. Weinheim: Juventa.
- Langer, M. F., Ziegele, F. & Hennig-Thurau, T. (2001). *Hochschulbindung. Entwicklung eines theoretischen Modells, empirische Überprüfung und Ableitung von Handlungsempfehlungen für die Hochschulpraxis*. Lehrstuhl Marketing I, Universität Hannover und CHE – Centrum für Hochschulentwicklung, Gütersloh. Zugriff am 03.07.2018. Verfügbar unter http://www.che.de/downloads/hs_bindung_bericht0105.pdf
- Lenecke, K. (2005). *Hochschulbindung durch Student Services. Grundlagen, Analyse, Perspektiven*. Saarbrücken: Verlag Dr. Müller.
- Marsh, H.W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In: Perry, R.P. & Smart, John C. (Hrsg.). *The scholarship of teaching and learning in higher education: An evidence-based perspective* (S. 319–384). New York: Springer.
- Meier, G. & Heckel, C. (2014). Online-Stichproben. In ADM-Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (Hrsg.). *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis* (S. 215–235). Wiesbaden: Springer VS.
- Nagy, G. (2005). *Berufliche Interessen, kognitive und fachgebundene Kompetenzen: Ihre Bedeutung für die Studienfachwahl und die Bewährung im Studium*. Dissertation, Freie Universität Berlin. Zugriff am 03.07.2018. Verfügbar unter <https://refubium.fu-berlin.de/handle/fub188/10012>
- Pfeiffer, H., Rach, H., Rosanowitsch, S., Wörl, J. & Schneider, M. (2015). Lehrevaluation. In: M. Schneider & M. Mustafić (Hrsg.). *Gute Hochschullehre: Eine evidenzbasierte Orientierungshilfe* (S. 153–184). Wiesbaden: Springer.
- Reis, O. & Ruschin, S. (2008). Kompetenzorientiert prüfen – Baustein eines gelungenen Paradigmenwechsels. In S. Dany, B. Szczyrba & J. Wildt (Hrsg.), *Prüfungen auf die Agenda! Hochschuldidaktische Perspektiven auf Reformen im Prüfungswesen* (S. 45–57). Bielefeld: Bertelsmann.
- Rindermann, H. (1997a). Die studentische Beurteilung von Lehrveranstaltungen. Forschungsstand und Implikationen für den Einsatz von Lehrevaluationen. In: R. S. Jäger, R. H. Lehmann & G. Trost (Hrsg.). *Tests und Trends 11* (S. 12–53). Weinheim: Beltz.
- Rindermann, H. (1997b). Generalisierbarkeit studentischer Veranstaltungsbeurteilungen. Sind Lehrevaluationsresultate generalisierbar auf andere Veranstaltungen eines Dozenten oder auf inhaltsgleiche Parallelveranstaltungen verschiedener Dozenten? In *Psychologie in Erziehung und Unterricht*, 44(3), 216–234.
- Rindermann, H. (1998). Das Münchner multifaktorielle Modell der Lehrveranstaltungsqualität. Entwicklung, Begründung und Überprüfung. In *Beiträge zur Hochschulforschung*, 3, 189–223.

- Rindermann, H. (1999). Bedingungs- und Effektivvariablen in der Lehrevaluationsforschung. Konzeption und Prüfung des Münchner multifaktoriellen Modells der Lehrveranstaltungsqualität. In *Unterrichtswissenschaft*, 27 (4), 357–380.
- Rindermann, H. (2001): *Lehrevaluation – Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen. Mit einem Beitrag zur Evaluation computerbasierter Unterrichts*. Landau: Empirische Pädagogik.
- Rindermann, H. (2003). Lehrevaluation an Hochschulen. Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. *Zeitschrift für Evaluation*, 2, 233–256.
- Rindermann, H. (2009). Qualitätsentwicklung in der Hochschullehre. In *Beiträge zur Lehrerbildung*, 27 (1), 64–73.
- Rindermann, H. (2016). Lehrveranstaltungsevaluation an Hochschulen. Der Einfluss der Rahmenbedingungen auf Qualität von Lehre und Ergebnisse von Lehrevaluation. In: Großmann, D. & Wolbring, T. (Hrsg.). *Evaluation von Studium und Lehre. Grundlagen, methodische Herausforderungen und Lösungsansätze* (S. 227–262). Wiesbaden: Springer VS.
- Rindermann, H. & Amelang, M. (1994). *Das Heidelberger Inventar zur Lehrveranstaltungsevaluation (HILVE)*. Handanweisung. Heidelberg: Asanger.
- Schiefele, U., Streblov, L. & Brinkmann, J. (2007). Aussteigen oder Durchhalten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39 (3), 127–140.
- Schneider, M. & Mustafić, M. (Hrsg.) (2015). *Gute Hochschullehre: Eine evidenzbasierte Orientierungshilfe*. Wiesbaden: Springer.
- Spady, W. G. (1970). Dropouts from Higher Education: An Interdisciplinary Review and Synthesis. *Interchange*, 1 (1), 64–85.
- Stichweh, R. (1979). Differenzierung der Wissenschaft. *Zeitschrift für Soziologie*, 8(1), 82–101.
- Wagner, P. & Hering, L. (2014). Online-Befragung. In Baur, N. & Blasius, J. (Hrsg.). *Handbuch Methoden der empirischen Sozialforschung* (S. 661–673). Wiesbaden: Springer VS
- Weinert, F. E. (1999). *Concepts of Competence. Contribution within the OECD project Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo)*. München: Max Planck Institute for Psychological Research.
- Wildt, J. (2009). Ausgelernt? Professor/innen im Prozess der Professionalisierung. *Organisationsberatung, Supervision, Coaching*, 16 (2), 220–227.

Autor/-innen

Dr. phil. Christian Herfter, Universität Leipzig, Allgemeine Didaktik und Schulpädagogik des Sekundarbereichs; Email: christian.herfter@uni-leipzig.de

Robert Aust, Universität Leipzig, Gleichstellungsbüro der Universität Leipzig; Email: robert.aust@uni-leipzig.de



Zitiervorschlag: Christian Herfter & Robert Aust (2018). Studentische Qualitätsurteile zu Ringvorlesungen. Überlegungen zu Theorie, Modell und Indikatoren. *die hochschullehre*, Jahrgang 5/2019, online unter: www.hochschullehre.org