

Reinmann, Gabi

Forschendes Lernen prüfen. Hochschuldidaktische Gedanken zu einer Theorie des Prüfens

Zeitschrift für Pädagogik 65 (2019) 4, S. 608-626



Quellenangabe/ Reference:

Reinmann, Gabi: Forschendes Lernen prüfen. Hochschuldidaktische Gedanken zu einer Theorie des Prüfens - In: Zeitschrift für Pädagogik 65 (2019) 4, S. 608-626 - URN: urn:nbn:de:0111-pedocs-239971 - DOI: 10.25656/01:23997

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-239971>

<https://doi.org/10.25656/01:23997>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipt.de
Internet: www.pedocs.de

ZEITSCHRIFT FÜR PÄDAGOGIK

Heft 4

Juli/August 2019

■ *Thementeil*

Kompetenzerwerb zukünftiger LehrerInnen in der universitären Ausbildung

■ *Allgemeiner Teil*

Der Einfluss von Lehrkräfteverbänden in der Steuerung
von Schulsystemen: Deutschland und Frankreich
im Vergleich

Zur Situationsspezifität des pädagogischen Ethos:
Eine empirische Studie im Bereich der betrieblichen
Berufsbildung

Forschendes Lernen prüfen: Hochschuldidaktische
Gedanken zu einer Theorie des Prüfens

Inhaltsverzeichnis

Thementeil: Kompetenzerwerb zukünftiger LehrerInnen in der universitären Ausbildung

<i>Olga Zlatkin-Troitschanskaia/Hans Anand Pant/Kai Cortina</i> Kompetenzerwerb zukünftiger LehrerInnen in der universitären Ausbildung. Einführung in den Thementeil	469
<i>Christoph Vogelsang/Andreas Borowski/David Buschhüter/ Patrick Enkrott/Maren Kempin/Christoph Kulgemeyer/Peter Reinhold/ Josef Riese/Horst Schecker/Jan Schröder</i> Entwicklung von Professionswissen und Unterrichtsperformanz im Lehramtsstudium Physik – Analysen zu valider Testwertinterpretation	473
<i>Sabrina Mathesius/Moritz Krell/Annette Upmeyer zu Belzen/Dirk Krüger</i> Überprüfung eines Tests zum wissenschaftlichen Denken unter Berücksichtigung des Validitätskriteriums relations-to-other-variables ...	492
<i>Colin Jeschke/Christiane Kuhn/Anke Lindmeier/ Olga Zlatkin-Troitschanskaia/Hannes Saas/Aiso Heinze</i> What is the Relationship Between Knowledge in Mathematics and Knowledge in Economics? Investigating the professional knowledge of (pre-service) teachers trained in two subjects	511
<i>Lara Pohle/Georg Hosoya/Catharina Loftfield/Lars Jenßen</i> Indicators Measuring Preschool Teachers' Stimulation Quality – Theoretical background and empirical testing	525
<i>Olga Kunina-Habenicht/Christina Maurer/Franziska Schulze-Stocker/ Kristin Wolf/Nora Hein/Detlev Leutner/Tina Seidel/Mareike Kunter</i> Zur curricularen Validität des BilWiss 2.0-Tests zur Erfassung des bildungswissenschaftlichen Wissens von (angehenden) Lehrkräften	542

Deutscher Bildungsserver

Linktipps zum Thema „Kompetenzerwerb zukünftiger LehrerInnen in der universitären Ausbildung“	557
--	-----

Allgemeiner Teil

Michael Dobbins/Rita Nikolai

Der Einfluss von Lehrkräfteverbänden in der Steuerung von Schulsystemen: Deutschland und Frankreich im Vergleich	564
---	-----

Sarah Forster-Heinzer

Zur Situationsspezifität des pädagogischen Ethos: Eine empirische Studie im Bereich der betrieblichen Berufsbildung	584
--	-----

Gabi Reinmann

Forschendes Lernen prüfen: Hochschuldidaktische Gedanken zu einer Theorie des Prüfens	608
--	-----

Dokumentation

Erziehungswissenschaftliche Habilitationen und Promotionen 2018	627
Impressum	U3

Table of Contents

Topic: Prospective Teachers' Competence Development in Higher Education

<i>Olga Zlatkin-Troitschanskaia/Hans Anand Pant/Kai Cortina</i> Prospective Teachers' Competence Development in Higher Education. An introduction	469
<i>Christoph Vogelsang/Andreas Borowski/David Buschhüter/ Patrick Enkrott/Maren Kempin/Christoph Kulgemeyer/Peter Reinhold/ Josef Riese/Horst Schecker/Jan Schröder</i> Development of Professional Knowledge and Teaching Skills in Academic Pre-service Physics Teacher Education: Validity analyses concerning the interpretation of test scores	473
<i>Sabrina Mathesius/Moritz Krell/Annette Upmeier zu Belzen/Dirk Krüger</i> An Evaluation of a Scientific Reasoning Test for Pre-service Biology Teachers: Using relations-to-other-variables as source of validity evidence	492
<i>Colin Jeschke/Christiane Kuhn/Anke Lindmeier/ Olga Zlatkin-Troitschanskaia/Hannes Saas/Aiso Heinze</i> What is the Relationship Between Knowledge in Mathematics and Knowledge in Economics? Investigating the professional knowledge of (pre-service) teachers trained in two subjects	511
<i>Lara Pohle/Georg Hosoya/Catharina Loftfield/Lars Jenßen</i> Indicators Measuring Preschool Teachers' Stimulation Quality – Theoretical background and empirical testing	525
<i>Olga Kunina-Habenicht/Christina Maurer/Franziska Schulze-Stocker/ Kristin Wolf/Nora Hein/Detlev Leutner/Tina Seidel/Mareike Kunter</i> On the Curricular Validity of the Bilwiss-2.0-Test for the Assessment of (Pre-service) Teachers' Educational Knowledge	542
<i>Deutscher Bildungsserver</i> Online Ressources "Prospective Teachers' Competence Development in Higher Education"	557

Articles

Michael Dobbins/Rita Nikolai

The Influence of Teachers' Unions on Output- and Competition-Oriented School Governance: Germany and France in comparison	564
---	-----

Sarah Forster-Heinzer

On the Situational Specificity of the Pedagogical Ethos: An empirical study in the field of vocational workplace training	584
--	-----

Gabi Reinmann

Assessing Research-Based Learning: Thoughts about a theory of assessment in higher education	608
---	-----

Habilitation Treatises and Dissertations in Education in 2018	627
---	-----

Impressum	U3
-----------------	----

Forschendes Lernen prüfen

Hochschuldidaktische Gedanken zu einer Theorie des Prüfens

Zusammenfassung: Trotz der Allgegenwart von Prüfungen an Hochschulen verfügen wir derzeit über keine genuin hochschuldidaktische Theorie des Prüfens. Dies macht sich ganz besonders in der Herausforderung bemerkbar, Leistungen im Zusammenhang mit dem Format des forschenden Lernens zu erfassen und zu bewerten. Der Beitrag beschreibt die damit verbundenen Widersprüche und Schwierigkeiten, erörtert Möglichkeiten des Prüfens analog zum Forschen und versucht, Impulse für die Erarbeitung einer theoretischen Klärung des akademischen Prüfens zu geben.

Schlagworte: Hochschullehre, Forschungsorientierung, Hochschulprüfungen, Prüfungsformen, Lernen durch Forschung

1. Ausgangslage

1.1 Spannungsfeld Hochschulprüfungen

„Die Prüfung gleicht [...] einer Wette, dass in der Zukunft etwas eintreten wird, was auch ohne die Prüfung eintreten würde, weil es eigentlich ohnehin schon gegenwärtig ist“ (Kaminski, 2011, S. 352). Eine solche Definition von Prüfung, wie sie der Technikphilosoph Andreas Kaminski vorschlägt, mag ungewöhnlich wirken. Doch es lohnt sich, über seine Aussage ein wenig nachzudenken, denn sie verrät ein paar wesentliche Dinge über das Prüfen: (a) Prüfungen liefern keine sicheren Ergebnisse. Wer prüft, arbeitet mit Annahmen, setzt also gewissermaßen auf etwas. (b) Das, worauf man setzt, liegt in der Zukunft. Dies gilt insbesondere dann, wenn sich Prüfungen an Kompetenzen ausrichten sollen, denn Kompetenzen sind bekanntermaßen Dispositionen zum Handeln – ein Vermögen also, Probleme in der Zukunft zu lösen. (c) Die Disposition selbst, so muss man voraussetzen, ist schon da – ganz unabhängig davon, ob sie geprüft wird oder nicht; wäre dem nicht so, ergäbe die Prüfung keinen Sinn. Was den Akteuren an Universitäten also relativ selbstverständlich vorkommt, weil es zum Alltag der Hochschullehre dazu gehört, nämlich zu prüfen und geprüft zu werden, ist am Ende ziemlich komplex. Was als Prüfungspraxis üblich, vielleicht sogar notwendig erscheint, darf dennoch hinterfragt werden, denn: Prüfungen sind historisch gewachsen; sie haben sich im Verlauf der Zeit immer wieder geändert: in ihrer Bedeutung, in ihrer Form, in ihrem Zweck, in ihrem Gegenstand, in ihrem Ausdruck der Bewertung wie Ziffern oder Umschreibungen (Nagel, 2017, S. 275; Ricken & Reh, 2017, S. 250). Es gilt, dies im Blick zu behalten: Sind nämlich Prüfungen in einer Gesellschaft nicht statisch, ist auch deren Zukunft offen.

Wer heute als Hochschullehrender Prüfungen plant und umsetzt, trifft auf etliche Erwartungen, mitunter gar in Form von Imperativen, die mit den vielfach bemühten Begriffen Qualität, Exzellenz und Diversität zu tun haben. Sie können das per se schon komplexe Prüfen zusätzlich erschweren, indem sie Lehrende mit Anforderungen konfrontieren, die rasch im Widerstreit liegen – ein Phänomen, auf das übrigens schon die Bundesassistentenkonferenz im Jahre 1970 hingewiesen hat.¹ So sind Hochschullehrende zum einen angehalten, eine hohe *Qualität* von Lehre und in der Folge auch eine solche von Prüfungen sicherzustellen. Qualität aber muss erst einmal bestimmt werden, wofür man (Qualitäts-)Standards setzt, damit man sich an diesen orientieren kann. Folgerichtig wird *Standardisierung* beim Prüfen zu einem Kriterium, das noch dazu gerecht erscheint: Wenn man sich auf Leistungsstandards im Studium einigt, also auf *sachliche* Bezugsnormen, wird die Prüfungspraxis für Studierende transparenter. Zum anderen sind Wissenschaftlerinnen und Hochschullehrende speziell in den letzten Jahren dazu aufgefordert, nicht nur in der Forschung, sondern ebenso in der Lehre *Exzellenz* anzustreben (vgl. z. B. Wissenschaftsrat, 2017). Exzellenz – auf welchen Ebenen auch immer – impliziert, sich mit anderen zu vergleichen und die Spitze zu erreichen, wofür man wiederum Rangreihen bildet. Zur Kohärenz der Argumentation passt, dass wir in der (exzellenten) Lehre auch die Besten finden und fördern (sollen). Zwangsläufig wird *Hierarchisierung* beim Prüfen zu einem Kriterium, und auch das erscheint in gewisser Weise gerecht: Setzt man auf *soziale* Bezugsnormen, werden diejenigen „belohnt“, die relativ zu den anderen die beste oder eine bessere Leistung zeigen. Schließlich sollen Hochschullehrende angesichts großer gesellschaftlicher Transformationsprozesse beim Lehren und folglich auch beim Prüfen sensibel für Diversität werden. *Diversität* markiert Unterschiede, die als gleichwertig anzuerkennen sind. Also müssen beim Prüfen genau diese Unterschiede berücksichtigt und die *Individualisierung* zu einem Kriterium gemacht werden. Wenn die *individuelle* Bezugsnorm angelegt wird, handelt man in dem Sinne gerecht, dass man dem Einzelnen gerecht werden will.

Prüfen, so ließe sich zusammenfassend folgern, ist komplex, persönlich herausfordernd, aber auch widersprüchlich. In Verbindung mit paradoxen Anforderungen sind Prüfungen eine ewige Quelle der Frustration bei allen Beteiligten: Prüfende Hochschullehrer sollen ohne wissenschaftlich belastbare Grundlagen von Performanzen auf Kompetenzen schließen und dabei Imperativen der Gerechtigkeit folgen, die kaum miteinander in Einklang zu bringen sind.

1 In der BAK-Schrift zum forschenden Lernen (1970) beschreibt Ludwig Huber unter anderem das Spannungsverhältnis zwischen Individualisierung und Objektivierung beim Prüfen.

1.2 Reaktionen der Hochschuldidaktik

Die Hochschuldidaktik hat, so meine Beobachtung, diesem hier nur knapp skizzierten Spannungsfeld nicht sehr viel entgegenzusetzen. Es erscheint erstaunlich, dass es keine *didaktische* Theorie des Prüfens gibt, oder wie es Norbert Ricken und Sabine Reh formulieren: dass das Prüfen bis heute eine Leerstelle pädagogischer Reflexion geblieben ist (Ricken & Reh, 2017, S. 247). Eingesprungen, um diese Leerstelle zu füllen, ist die Psychologie. Bereits um 1900 hat sie das Prüfungswesen unserer Gesellschaft wissenschaftlich und praktisch enorm beeinflusst (Kaminski, 2011): Auslese- und Eignungsprüfungen, Intelligenz- und Willensprüfungen in Schule, Wirtschaft und Militär haben bereits vor über hundert Jahren die Pflöcke für unser heutiges Prüfungswesen eingeschlagen. Auch an Universitäten ist es kaum möglich, sich der Psychologisierung des Prüfens zu entziehen (Alexander, 2018, S. 148–149). Während daher Pädagogen heute eine wissenschaftliche Leerstelle beim Prüfen identifizieren, sprechen Psychologen wie Richard Mayer (2018) von einer „Science of Assessment“. Sie meinen damit eine an den Naturwissenschaften orientierte *psychologische* Theorie des Prüfens bzw. eine test- und messtheoretische Diagnostik. Unter dem Stichwort des kompetenzorientierten Prüfens findet diese, zumindest theoretisch, auch Einzug in die Hochschullehre. Würde man aber tatsächlich, etwa analog zu Large Scale Assessments wie PISA im Schulbereich, eine Kompetenzdiagnostik an der Hochschule praktizieren wollen, müsste man auch so vorgehen, und das hieße sehr vereinfacht formuliert (vgl. z. B. Wolter & Schiener, 2016): Wir bräuchten für jeden Studiengang theoretische Kompetenzstrukturmodelle; die Modelle müssten wir in Testaufgaben übersetzen; die Aufgaben wären zunächst von Probanden zu bearbeiten; die resultierenden empirischen Daten müssten wir mit statistischen Verfahren auswerten und Kompetenzindikatoren festlegen; schließlich bräuchten wir noch Kompetenzniveaumodelle, mit denen wir Kompetenzstufen definieren. Dann, und erst dann, könnten wir diese Kompetenztests als Prüfungsverfahren einsetzen. Dass das weit weg von der Praxis ist, leuchtet rasch ein. Es handelt sich hier um Kompetenzforschung, nicht um Prüfungspraxis.

Trotzdem sind mess- und testtheoretische Ansätze die implizite, manchmal auch explizite Referenz selbst von vereinfachten Verfahren und Empfehlungen für das Prüfen, wie sie etwa im Zuge des *Constructive Alignments* hochschuldidaktisch nahegelegt werden (vgl. Reeves, 2006; Iller & Wick, 2009; Schaper & Hilkenmeier, 2013; Klauer & Leutner, 2014, S. 201). Empfohlen wird hier, mit der Definition eines Lehrziels, orientiert an psychologischen Lehrzieltaxonomien, zu beginnen, dazu Aufgaben mit unterschiedlichen Lösungswahrscheinlichkeiten für die Prüfung zu konstruieren und die Lehraktivitäten darauf auszurichten. Das klingt vernünftig und plausibel, auch wenn inzwischen kritische Analysen einer durchgängigen Praxis am Prinzip des *Constructive Alignments* vorliegen (vgl. Murtonen, Gruber & Lethinen, 2017; Reinmann, 2018). Die derzeit weite Verbreitung und zunehmende Akzeptanz des *Constructive Alignment* haben zudem den erfreulichen Nebeneffekt, dass Hochschullehrende nach neuen Prüfungsformen suchen: ein Bedarf, der auch gedeckt wird. So ist etwa – um nur ein aktuelles Beispiel zu nennen – kürzlich ein Buch erschienen mit 53 Prüfungsformaten für die

Hochschullehre und dem Versprechen, die studentischen Kompetenzen nun auch kompetent zu prüfen (vgl. Gerick, Sommer & Zimmermann, 2018). Dort findet sich Traditionelles, Alternatives und Phantasievolles, also zum Beispiel: traditionelle Formate wie mündliche Prüfung, Hausarbeit und Klausur, sogenannte alternative Formate wie Poster und Portfolio sowie phantasievoll bezeichnete Spezialformate wie Pecha Kucha, Expertenboard oder Coding Challenge.

Sammlungen dieser Art, so praktisch wertvoll sie auch sein mögen, zeugen von einem gewissen Aktionismus: Bisherige Formen des Prüfens werden aussortiert oder verändert; neue Prüfungsformen werden ausprobiert, verschwinden wieder oder etablieren sich. Digitale Technologien beflügeln manche Ideen im Reich des Prüfens, Prüfungsordnungen verweisen sie in ihre Schranken. So ermutigend das insgesamt sein mag, so ernüchternd ist doch, dass man, obschon auf universitärem Boden stehend, beim Prüfen didaktisch gesehen eher unsystematisch agiert: Einerseits beugt man sich zunehmend und häufig ohne erkennbare Not den Vorgaben aus der psychologischen Kompetenzdiagnostik. Andererseits erliegen viele Hochschullehrende immer häufiger und ohne theoretische Reflexion dem Druck, nun auch beim Prüfen möglichst innovativ zu sein.

1.3 Zwischenfazit und Überblick

Eine hochschuldidaktische Theorie des Prüfens liegt derzeit nicht vor, dafür widersprüchliche Erwartungen, die sich in widerstreitenden Bezugsnormen niederschlagen. Psychologische Annahmen und Befunde aus Testtheorie und Kompetenzdiagnostik stopfen die theoretische Lücke, werden zwar praktisch kaum wirksam, steuern aber trotzdem mehr oder weniger implizit das Denken über das Prüfen in der Hochschullehre. Genau genommen weiß man nicht viel über das akademische Prüfen und Geprüft-Werden, ist aber dennoch – aus praktischen Gründen – ständig damit beschäftigt und sucht nach Innovationen, ohne dafür eine didaktische oder pädagogische Theorie zu haben. Diese komplexe Ausgangssituation beim akademischen Prüfen wird nun ganz besonders deutlich, wenn es darum geht, Prüfungen im Kontext des *forschenden Lernens* zu gestalten. Denn Lehre, die Studierende darin fördert, selbst forschend tätig zu werden, steht wie keine andere Form der Lehre für die Besonderheit universitärer Bildung. Genau hier müsste sich denn auch eine hochschuldidaktische Theorie des Prüfens beweisen.

Im Folgenden versuche ich, mich einer hochschuldidaktischen Theorie des Prüfens *anzunähern*, indem ich die Möglichkeiten des Prüfens beim forschenden Lernen ausleuchte. Dazu gehe ich in drei Schritten vor: Zunächst gebe ich einen kurzen Überblick über wesentliche Merkmale und Abgrenzungsmöglichkeiten forschenden Lernens zu anderen, verwandten, Lernformen und gehe knapp auf die Varianten innerhalb des Konzepts ein. Meine Argumentation hin zu einer hochschuldidaktischen Theorie des Prüfens beginne ich mit der Vorstellung von zwei Systematiken, mit denen man Prüfungsformate theoriegeleitet generieren kann. Systematische Modellierungen sichtbarer (bzw.

manifesten) Erscheinungen des Prüfens könnten nämlich *eine* Vorstufe zu einer didaktischen Prüfungstheorie sein. Im Anschluss wende ich mich den eher unsichtbaren (bzw. für die Praxis oft latenten) Funktionen und Strukturen des Prüfens inklusive Fragen der Kohärenzbildung zu, die für eine hochschuldidaktische Prüfungstheorie mindestens ebenso wichtig sein dürften wie die Frage nach den Prüfungsformaten.²

2. Forschendes Lernen als Gegenstand akademischen Prüfens

2.1 Definition und Abgrenzung

In den letzten Jahren sind zahlreiche Bücher erschienen, die sich dem forschenden Lernen widmen; dazu kommen viele Artikel in Handbüchern und Zeitschriften sowie unzählige Initiativen und Projekte zum forschenden Lernen an unseren Hochschulen.³ Das Konzept hat also offenbar eine hohe Attraktivität. Allerdings wird der Begriff des forschenden Lernens relativ großzügig, mitunter auch inflationär (und entsprechend unpassend) verwendet, obschon es seit langem Definitionsversuche gibt, die sich im Kern durchaus ähneln. Neben der aus dem Jahr 1970 stammenden Definition forschenden Lernens von der Bundesassistentenkonferenz in der noch heute viel zitierten programmatischen Schrift zum forschenden Lernen und wissenschaftlichen Prüfen (BAK, 1970) wird wohl am häufigsten die von Ludwig Huber aus dem Jahr 2009 herangezogen:

Forschendes Lernen zeichnet sich vor anderen Lernformen dadurch aus, dass die Lernenden den Prozess eines Forschungsvorhabens, das auf die Gewinnung von auch für Dritte interessanten Erkenntnissen gerichtet ist, in seinen wesentlichen Phasen – von der Entwicklung der Fragen und Hypothesen über die Wahl und Ausführung der Methoden bis zur Prüfung und Darstellung der Ergebnisse in selbstständiger Arbeit oder in aktiver Mitarbeit in einem übergreifenden Projekt – (mit) gestalten, erfahren und reflektieren. (Huber, 2009, S. 11)⁴

Die in der Definition erwähnten „anderen Lernformen“, von denen man das forschende Lernen abgrenzen kann und soll, sind durchaus zahlreich. Exemplarisch mögen an dieser Stelle drei relevante Abgrenzungen genügen, nämlich zum (a) problemorientierten Lernen, (b) Projektstudium und (c) wissenschaftlichen: (a) Beim forschendem Lernen

-
- 2 In der weiteren Behandlung des Themas müsste die Frage der Lehr-Lernziele/-ergebnisse (z. B. Kompetenzen) des hier fokussierten forschenden Lernens einbezogen werden (vgl. Huber & Reinmann, in Vorbereitung).
 - 3 Diese sind vielerorts dem Qualitätspakt Lehre geschuldet, also mit zusätzlichen Ressourcen vom Bund finanziert, ohne dass bislang klar ist, wie es im Jahr 2021 damit weitergeht, wenn die Förderungen enden.
 - 4 Die beiden Definitionen weichen nur geringfügig voneinander ab, was wohl auch darin begründet liegt, dass Ludwig Huber an der damaligen BAK-Schrift (2009 neu aufgelegt) beteiligt war.

entwickeln Studierende eine Frage- oder Problemstellung, die sie interessiert, oder sie entscheiden sich, eine solche vom Lehrenden zu übernehmen, und machen sich diese zu eigen. Die Frage- oder Problemstellung ist so gewählt, dass Ergebnisse möglich sind, die nicht nur nicht nur als Lerngewinn für die Studierenden selbst, sondern auch *für Dritte* von Interesse sein könnten und daher auch anderen mitgeteilt werden sollten. Erst infolge der Mitteilung von Erkenntnissen entsteht Wissenschaft und werden forschend Lernende Teil der wissenschaftlichen Gemeinschaft (vgl. von Hentig, 1970). Forschendes Lernen mag also an Problemen orientiert und damit dem problemorientierten Lernen in dieser Hinsicht ähnlich sein. Problemorientiertes Lernen aber läuft darauf hinaus, dass sich Studierende zwar selbständig Methoden und Inhalte erarbeiten, dies aber – genau anders als beim forschenden Lernen – von *vorgegebenen* Fragen aus tun, die zu Ergebnissen führen, die im Prinzip *bekannt* sind (vgl. Reinmann, 2016a). (b) Forschung vollzieht sich meistens in Projekten; wir sprechen denn auch ganz selbstverständlich von Forschungsprojekten. Nicht jedes Projekt aber ist ein Forschungsprojekt – auch an Hochschulen nicht. Projektorientierung ist ein didaktisches Prinzip mit vielen Gesichtern; in den 1970 Jahren machte es als Projektstudium Karriere – damals allerdings als kritisches Konzept und ausdrücklich *gegen* forschendes Lernen: Projekte im Studium sollten nicht primär wissenschaftlichen Interessen geschuldet sein und vorrangig zu neuen Erkenntnissen führen (vgl. Huber, 2013). Sie sollten ein gesellschaftlich relevantes Problem in kritischer Absicht aufgreifen *und* gesellschaftliche Praxis verändern. Das ist nach wie vor eine wichtige Unterscheidung etwa zwischen forschendem Lernen und dem heute ebenfalls viel genutzt didaktischen Konzept des *Service Learning* – auch wenn Forschungsergebnisse natürlich eine Wirkung mit gesellschaftlicher Relevanz haben können, was aber beim forschenden Lernen weder zwingend noch der Ausgangspunkt ist. (c) Forschendes Lernen vollzieht sich im Medium der Wissenschaft, wobei man natürlich während eines akademischen Studiums immer erwarten dürfte, dass wissenschaftlich gearbeitet wird. Zum wissenschaftlichen Arbeiten zählt man in der Regel alles, was nach fachspezifischen Gepflogenheiten als logisch, systematisch, methodisch und nachprüfbar gilt, was Studierende für Studien- oder Prüfungsarbeiten recherchieren, vortragen, aufschreiben usw. Wissenschaftliches Arbeiten orientiert sich an vorhandenem Wissen, das unter einer bestimmten Frage- oder Problemstellung zusammengestellt, aufbereitet und bewertet wird. Es ist je nach Fach vom forschenden Lernen mitunter schwer abzugrenzen und in jedem Fall stets Teil forschendes Lernens. Dieses aber impliziert, dass *alle* Phasen eines Forschungsprozesses durchlaufen werden und Studierende den *ganzen* Bogen des Forschens kognitiv, emotional und sozial erfahren.

2.2 Varianten

Forschendes Lernen im eben beschriebenen Sinne ist nur *eine* Variante, wie man Forschen und Lernen durch die Gestaltung von Hochschullehre in ein Verhältnis setzen kann. Andere Varianten enthalten nicht alle definitorischen Elemente des forschenden Lernens im engeren Sinne, werden aber mitunter dennoch als forschendes Lernen

bezeichnet oder erhalten andere, ähnlich klingende, Namen.⁵ Diese begriffliche Unschärfe ist für Theoriebildung und empirische Studien gleichermaßen problematisch. Es scheint sich – mit leichter Variation – inzwischen durchzusetzen, nicht nur, aber vor allem drei Typen forschungsnahen Lernens zu unterscheiden (vgl. Huber, 2014; Reinmann, 2016b):

- Das eben skizzierte und in diesem Text im Fokus stehende forschende Lernen (im engeren Sinne) ließe sich als *Lernen durch Forschung* kennzeichnen. Es dürfte ange raten sein, als *forschendes* Lernen nur diese Form des Lernens zu benennen, bei der Studierende tatsächlich selbst forschen (unabhängig davon, ob sie das völlig eigenständig, begleitet oder phasenweise auch angeleitet tun).
- Nicht immer aber können, wollen oder sollten Studierende alle Phasen eines Forschungszyklus durchlaufen. Man kann den Prozess des Forschens im Studium auch in Ausschnitten oder probenhalber umsetzen, sich in einzelnen Forschungstätigkeiten versuchen und dabei erkennen, was Forschung ausmacht. Bei einem solchen *Lernen für die Forschung* hat die Beziehung zwischen Lernen und Forschung eher übenden Charakter, richtet sich aber am Prozess der Forschung aus (*forschungsorientiertes Lernen*).
- Schon die BAK (1970) hat ein als genetisch charakterisiertes Lernen in Erwägung gezogen, um Lernen und Forschung miteinander in Beziehung zu setzen. Studierende, so der Grundgedanke, sollen Forschung mindestens nachvollziehen, was allerdings nicht über einfache Wissensvermittlung funktioniert, sondern Gelegenheiten bedarf, den Weg nachzuverfolgen, wie etwas wissenschaftlich fragwürdig, also eine Frage zur Forschung wird, und was die wissenschaftliche Generierung von Wissen ausmacht. Ein solches *Lernen über Forschung* fußt auf der Forschung, ohne selbst forschend zu werden (*forschungsbasiertes Lernen*).

2.3 Zwischenfazit

Es mag schwerfallen, übendes Lernen für die Forschung vom forschenden Lernen abzugrenzen; weniger schwierig dürfte es sein, die Unterschiede zwischen forschendem Lernen und nachvollziehendem Lernen über Forschung zu erkennen. Alle Varianten der Beziehung zwischen Lernen und Forschung haben in der akademischen Lehre ihre Berechtigung. Dass der vorliegende Beitrag das forschende Lernen im engeren, hier definierten und abgegrenzten, Sinne ins Zentrum stellt, impliziert keine Bewertung oder Empfehlung, nur mehr auf forschendes Lernen zu setzen. Vielmehr lassen sich an diesem, in seiner Attraktivität offenbar wachsenden, anspruchsvollen Konzept besonders gut die *akademischen* Anforderungen an das Prüfen innerhalb der eingangs skizzierten komplexen Ausgangslage an unseren Hochschulen durchdenken. Dazu kommt, dass die

5 Das gilt sowohl für die deutsche als auch für die englische Fachliteratur.

zunehmende Verbreitung forschenden Lernens die, im Vergleich zu anderen Varianten forschungsnahen Lernens, größte Herausforderung für das Prüfen mit sich bringt, die bislang nicht systematisch angegangen wird.

Im Folgenden gilt es, sich zunächst ein Bild über die bestehende Prüfungspraxis an Hochschulen zu machen, denn: Auch das Prüfen forschenden Lernens muss *anschlussfähig* an das Bestehende sein. Gleichzeitig sollte es jedoch *spezifisch* für die Forschung im Prozess des Lernens sein, worauf sich die zweite Systematik stützt und entsprechend Anleihen bei Leistungsnachweisen in der Forschung nimmt.

3. Modellierung von manifesten Erscheinungen des Prüfens

3.1 Generische Prüfungsformen auf Basis der Prüfungspraxis

Ziel dieses Abschnitts ist es, eine überschaubare Anzahl von Grundformen des Prüfens festzulegen, die sich vielseitig ausdifferenzieren lassen. Ich spreche daher von generischen Prüfungsformen, mit denen man (weitere) vielfältige Prüfungsvarianten kreieren kann. Mein Ausgangspunkt ist die verbreitete Einteilung in mündliche, schriftliche und praktische Prüfungsformen. In Anlehnung an Jerome Bruners Repräsentationsmodi⁶ könnte man mündliche und schriftliche Prüfungen als symbolische Formen bezeichnen und praktische Prüfungen als enaktive Formen. Die drei Prüfungsformen lassen sich relativ trennscharf in je zwei Unterformen aufteilen:

- 1) Mündliche Prüfungsformen sind eher monologisch oder dialogisch ausgerichtet. In monologischen Prüfungen referieren oder präsentieren Studierende das, was sie gelernt oder erarbeitet haben. Man könnte sich auf *Vortrag* als Bezeichnung für diese generische Form des Prüfens einigen. In dialogischen Prüfungen wird die Beziehung interaktiver und es werden Fragen gestellt. Um wieder ein neutrales Wort zu wählen, könnte man diese generische Form des Prüfens als *Gespräch* bezeichnen.
- 2) Schriftliche Prüfungen lassen sich relativ gut und deutlich sichtbar darin unterscheiden, ob die zu Prüfenden ihre Leistung erbringen, wenn sie an der Universität präsent sind, oder ob sie abwesend sein können, weil die Prüfungsleistung nicht vor Ort erbracht werden muss. Der typische Fall für eine schriftliche Prüfung in Präsenz ist die *Klausur* – ein Begriff, den alle Disziplinen verstehen dürften. Schriftliche Prüfungen in Absenz dagegen verdichten sich wohl am ehesten im bekannten Terminus der *Hausarbeit*.
- 3) Praktische bzw. enaktive Prüfungen sind je nach Disziplin unterschiedlich verbreitet, oft aufwändig und selten einheitlich dargestellt. Das merkt man spätestens dann, wenn man versucht, diese Formen des Prüfens möglichst neutral zu beschreiben – es fehlen einem schnell die Worte. Ich schlage vor, situierte Prüfungen von material-

6 Erstmals in Bruner, Olver & Greenfield (1966).

sierten Prüfungen zu unterscheiden. Situieret sind solche Prüfungen, in denen die zu Prüfenden etwas zeigen, weshalb sie sich auch als *Demonstration* bezeichnen lassen. Materialisiert sind solche Prüfungen, in denen man das Ergebnis eines praktischen Tuns bewertet. Man könnte diese Prüfungsform als eine *Produktion* ansehen.

Generisch sind die Grundformen des Prüfens (Vortrag, Gespräch, Klausur, Hausarbeit, Demonstration, Produktion), weil sich damit beliebig viele Prüfungsvarianten kreieren (und kombinieren) lassen. Die Gestaltungsdimensionen, auf denen man die Grundformen modellieren kann, sind vielfältig. Exemplarisch seien folgende Dimensionen genannt:

- a) *Sozialform*: Alle genannten Grundformen des Prüfens können z.B. einzeln oder in Gruppen stattfinden.
- b) *Ressourcen*: In mündlichen und schriftlichen Grundformen des Prüfens können Hilfsmittel und Unterlagen zugelassen, verbannt oder eingeschränkt nutzbar sein.
- c) *Umsetzungsbedingungen*: Für fast alle Grundformen des Prüfens sind z.B. Feld- oder Laborumstände denkbar, authentische oder künstliche Settings denkbar.
- d) *Einsatz digitaler Medien*: Alle sechs Grundformen des Prüfens lassen sich daraufhin analysieren, inwiefern sie digital unterstützt oder digital umgesetzt werden können.
- e) *Zweck*: Beim Prüfen können die zu Prüfenden Wissen wiedergeben, anwenden oder auch neu erschaffen. In einer Hausarbeit etwa kann man Bestehendes referieren, ein gegebenes Problem mit bekanntem Wissen lösen oder eigene Ergebnisse darstellen und reflektieren.

Im Prinzip lassen sich die genannten generischen Prüfungsformen allesamt beim forschenden Lernen anwenden und forschungsbezogen ausgestalten: Studierende können *vortragen*, wie sie geforscht und was sie herausgefunden haben. Sie können daran anschließend oder auf Basis eines Posters ihre Forschung im *Gespräch* verteidigen und Fragen beantworten. Studierende können selbstredend über ihre Forschung in einer *Hausarbeit* berichten. *Klausuren* erscheinen dann möglich, wenn Studierende darin Wissen anwenden können, das sie im Forschungsprozess erworben haben, etwa Methodenwissen an einem Fall. Studierende können *demonstrieren*, wie sie bestimmte Forschungsmethoden beherrschen. Dazu müsste man sie beobachten, etwa wie sie ein Experiment durchführen, jemanden interviewen, Daten analysieren und ähnliches. Schließlich kann man diejenigen Artefakte zur Prüfung heranziehen, die im Prozess des Forschens direkt *produziert* werden. Ein Teil dieser Artefakte wird schriftlicher Natur sein, z.B. Literaturübersichten zum Forschungsgegenstand oder die Zusammenstellung von Ergebnissen. Prinzipiell aber sind alle Materialisierungsformen denkbar, etwa neu entwickelte Forschungsinstrumente, generierte Rohdaten, entdeckte Sachverhalte, konstruierte Innovationen etc. (vgl. Reinmann, 2017).

3.2 Analoge Prüfungsformen zu Leistungsnachweisen in der Forschung

Während die vorgeschlagenen generischen Prüfungsformen in ihrer Systematik unabhängig sind von möglichen didaktischen Konzepten (und somit nicht nur auf das forschende Lernen, sondern auch auf die verwandten Formen des forschungsorientierten und forschungsbasierten Lernens hin ausgestaltet werden können), ergibt eine direkte Ausrichtung des Prüfens am Forschen vorzugsweise beim forschenden Lernen einen Sinn. Hier liegt es zunächst einmal nahe, danach zu fragen, ob es im Forschungsprozess selbst so etwas wie Nachweise für Forschungsleistungen gibt, die man als Leistungsnachweise in der Lehre heranziehen könnte, die das forschende Lernen fördern will. Eine erste Antwort darauf haben die Hochschuldidaktiker Peter Tremp und Thomas Hildbrand vor einigen Jahren im Zürcher Framework formuliert (vgl. Tremp & Hildbrand, 2012). Ich möchte den Grundgedanken des Modells aufgreifen und weiterdenken.

Das Zürcher Framework ordnet dem Forschungsprozess nicht nur übliche Lehrformate zu, sondern auch Prüfungsformate. Gemeint ist nicht, alle Zwischenschritte in der Forschung einer Prüfung zu unterziehen. Es geht darum, Prüfungsformen, die ein Pendant in einem Forschungsprozess haben, bevorzugt zu nutzen. Chronologisch betrachtet sind das z. B. das Thesenpapier, in dem eine Forschungsfrage entwickelt wird, die Forschungsübersicht, in der eine Befundlage gesichtet wird, das Laborjournal, in dem der Ablauf einer Untersuchung beschrieben wird, der Ergebnisbericht usw. Nun könnte man auch unabhängig von der im Zürcher Framework favorisierten Chronologie im Forschungsprozess noch etwas genauer darauf schauen, was denn eigentlich die *Leistungsnachweise* in der Forschung sind: Woran macht sich die Leistung – gegebenenfalls auch die Qualität, Exzellenz oder Originalität – einer Forschungsarbeit fest? Zu denken wäre da zunächst an die unmittelbar resultierenden Ergebnisse eines Forschungsprozesses, die je nach Typus der Forschung ausgesprochen vielfältig sein können: ein gesättigtes Argument, ein überprüftes Modell, Rohdaten aus einer Erhebung, ein technisches Konstrukt, eine entdeckte Substanz etc. Unmittelbare Ergebnisse dieser Art sind prinzipiell dazu geeignet, etwas über die Güte einer Forschung auszusagen, sodass sie zu Prüfungszwecken beim forschenden Lernen nutzbar wären; ich nenne sie *unmittelbare Nachweise der Forschungsleistung*. Wenn wir uns heute in der Forschungspraxis vermehrt mit dem Thema Forschungsdatenmanagement auseinandersetzen müssen, dann bewegen wir uns auf einer vergleichbaren Abstraktionsebene. Unmittelbare Nachweise der Forschungsleistung kommen dem nahe, was ich bei den generischen Prüfungsformen als Materialisierung aus der Forschung bezeichnet habe.

In aller Regel allerdings kommunizieren Wissenschaftlerinnen unmittelbare Nachweise der Forschungsleistung nicht direkt, sondern wählen die Veröffentlichung als Kommunikationskanal in die wissenschaftliche Gemeinschaft. Mit anderen Worten: Sie verwenden das geschriebene oder gesprochene Wort, um Ergebnisse der Forschung zu vermitteln – in der Regel am Ende, mitunter auch im Prozess, sofern Zwischenergebnisse mitteilungswürdig sind. Veröffentlichen kann man Erkenntnisse in Monografien, Buchbeiträgen und Zeitschriftenartikeln ebenso wie in mündlichen Beiträgen und Postern auf Konferenzen. Hier liegt eine hoch analoge Struktur zur bestehenden Prüfungs-

praxis in der Hochschullehre vor: Man denke nur an mündliche monologische sowie an schriftliche Prüfungsformen in Absenz. In der Logik der Forschung erscheint es naheliegend, von *vermittelten Nachweisen der Forschungsleistung* zu sprechen, die sich auch zur Prüfung forschenden Lernens eignen.

In der heutigen Forschungspraxis werden zunehmend Indikatoren dominant, wenn es darum geht, Forschungsleistungen nachzuweisen und zu bewerten (vgl. Bornmann & Marx, 2014). Das heißt: Nicht die unmittelbaren Forschungsergebnisse und auch nicht deren verbale Vermittlung werden zur Einschätzung der Leistung herangezogen, sondern aggregierte Maße wie der Impact Factor von Publikationsorganen sowie Zitationsindizes. Hier hat man es mit einer Form von Leistungsnachweisen in der Forschung zu tun, die nicht nur Komplexität reduziert und entsprechende Vorteile hat, sondern auch gravierende Risiken mit sich bringt: Man konzentriert sich auf Kennzahlen, vertraut auf deren Validität und beschäftigt sich nicht mehr mit dem eigentlichen Inhalt der Forschung (vgl. Mau, 2017). Es handelt sich hier um *metrische Nachweise der Forschungsleistung*. Für die Prüfungspraxis an Hochschulen sind sie wohl kaum brauchbar: Studentische Leistungen gelangen zu selten auf diesen „Markt“ und es spricht nicht so viel dafür, einen solchen auch noch nachzubilden.

3.3 Zwischenfazit

Auf dem Weg zu einer didaktischen Prüfungstheorie – exemplarisch durchdacht am forschenden Lernen – kann es sinnvoll sein, sich zunächst mit Prüfungen als Erscheinungsformen zu beschäftigen und dafür klare Beschreibungsdimensionen zu finden: als Alternative zu langen Listen mit Formaten, die man schlecht vergleichen kann, und zu willkürlich gewählten Bezeichnungen, die sich kaum dechiffrieren lassen. Ich habe dazu zwei Systematiken angeboten: Die erste stützt sich auf generische Prüfungsformen, die an der gegenwärtigen Prüfungspraxis ansetzt. Die zweite orientiert sich analog an Leistungsnachweisen im Forschungsprozess. Beide Systematiken sind kompatibel und kombinierbar. Für das forschende Lernen ließen sich unmittelbare und vermittelte Nachweise der Forschungsleistung als Prüfungsformen nutzen, die man mithilfe der sechs generischen Prüfungsformen weiter fachspezifisch ausgestalten könnte.

Was man beobachtet, wenn es um das Prüfen geht, ist das eine; ich habe diese Beobachtungen manifeste Erscheinungen genannt. Welche Funktionen und Strukturen das Prüfen hat, ist das andere; da sich Funktionen und Strukturen nicht selten der direkten Beobachtung entziehen, spreche ich von latenten Bedeutungen, die daraus für die beteiligten Akteure entstehen. Ordnet und kreiert man Prüfungsformate, bewegt man sich auf der Ebene der manifesten Erscheinungen von Prüfungen. Denkt man darüber nach, wozu geprüft wird und was das Prüfen letztlich ausmacht, befindet man sich auf der Ebene der (eher) latenten Bedeutungen des Prüfens.

4. Auseinandersetzung mit latenten Bedeutungen des Prüfens

4.1 Funktionen des Prüfens

Üblicherweise werden zwei große Gruppen von Funktionen unterschieden: gesellschaftliche und didaktische (Macke, Hanke, Viehmann-Schweizer & Raether, 2016, S. 141–143). Genauer müsste es heißen: Es kommt darauf an, aus welcher *Perspektive* man Prüfungen betrachtet. Nimmt man die *gesellschaftliche* Perspektive ein, dann lautet die Frage: Was bringen universitäre Prüfungen den Anschlusssystemen der Universität und den Bürgern, die eine Gesellschaft ausmachen? Nimmt man eine *didaktische* Perspektive ein, dann ist die Frage: Was bringen universitäre Prüfungen den Akteuren, die am didaktischen Geschehen direkt beteiligt sind, also den Studierenden und Lehrenden?

Aus der Perspektive der *Gesellschaft* erwartet man von universitären Prüfungen gemeinhin, dass sie einen Beitrag zu folgenden drei Aufgaben leisten:

- erstens die richtigen Personen an die richtigen Stellen in der Gesellschaft bringen – das ist eine Auslese- und Rekrutierungsfunktion;
- zweitens verlässliche Aussagen über die Leistung und das Leistungspotenzial, sprich Kompetenzen, von Personen machen – das ist eine Diagnose- und Prognosefunktion; und
- drittens die Qualität der Universität als forschungsbezogene Bildungseinrichtung sicherstellen – das ist eine Kontroll- und Legitimationsfunktion.

Aus der Perspektive der *Didaktik* erwartet man von universitären Prüfungen, dass sie dazu beitragen, das Lernen und Lehren im weitesten Sinne zu verbessern und einen pädagogischen Auftrag zu erfüllen. Es ist üblich, hier zwischen formativen und summativen Funktionen zu unterscheiden. In den dazu gängigen englischen Bezeichnungen wird das besser greifbar:

- Formative Prüfungen entsprechen dem *Assessment for Learning*. Damit wird deutlich, dass es sich um – in der Regel kleinere – Leistungsnachweise handelt, die im *Prozess* des Lernens dazu dienen, Studierende und Lehrende über den aktuellen Stand des Wissens und Könnens zu informieren. Die Prüfung ist hier *für* das Lernen da.
- Summative Prüfungen entsprechend dem *Assessment of Learning*. Das bringt zum Ausdruck, dass es sich um – in der Regel umfangreichere – Leistungsnachweise handelt, die am Ende eines Lernprozesses dazu dienen, *Ergebnisse*, also den erreichten Stand des Wissens und Könnens von Studierenden, zu ermitteln (der unter anderem dazu verwendet werden kann, Rückschlüsse über die Qualität des Lehrens zu ziehen). Die Prüfung macht hier eine Aussage *über* das Lernen.

Vielerorts werden formative Funktionen mit unbenotetem Prüfen assoziiert und summative Funktionen mit Benotung (Schaper & Hilkenmeier, 2013, S. 28–29). Das ist zwar

plausibel – jedenfalls unter der Annahme, dass ein *Assessment for Learning* durch Noten gestört wird und ein *Assessment of Learning* ohne Noten keine Wirkung hat. Beides aber ist nicht zwingend: Nicht immer ist leicht zu beurteilen, ob Prüfungen noch Einfluss auf den laufenden Lernprozess nehmen und formativ wirken oder abschließenden Charakter haben und summativ sind. Noten können Studierende sowohl motivieren als auch Stress verursachen. Summative Prüfungen setzen meist auf Noten, können aber prinzipiell auch anders bewertet werden.

Natürlich aber ist die Frage der *Benotung* entscheidend für den Charakter und das Erleben von Prüfungen. Noten werden in der Regel dann vergeben, wenn Prüfungen Rechtsfolgen haben⁷, wenn sie also darüber mitentscheiden, wie sich der Verlauf des Studiums gestaltet oder das Gesamtergebnis aussieht. Rechtsfolgen haben Noten in der Regel tatsächlich nur bei summativen Prüfungen. Diese Art des Prüfens fungiert dann in der Regel auch als Grundlage für die gesellschaftlich erwartete Auslese, Rekrutierung, Diagnose, Prognose, Kontrolle und Legitimation. Genau das fixiert die Prüfungspraxis an Universitäten vermutlich stark auf das summative Prüfen, während die didaktische Funktion des formativen Prüfens mitunter gar nicht mehr als Prüfung interpretiert wird. Das war nicht immer so: Geht man auf die lateinische Wurzel des Wortes Prüfen, zurück, nämlich *probare*, trifft man einen recht weiten Bedeutungsrahmen, der das Probieren und Erproben ebenso umfasst wie das Bemessen und Beurteilen (Ricken & Reh, 2017, S. 250). Das heutige Verständnis von Prüfung hat sich auf Letzteres deutlich verengt.

Was folgt daraus für das Prüfen beim forschenden Lernen? Einer Antwort kommt man wohl nur näher, wenn man auch hier nach analogen Bezügen in der Forschung sucht. Und in der Tat: Zwischen Forschungs- und Lehrpraxis finden sich Parallelen in der Funktion von Leistungsnachweisen: Vor allem metrische, aber auch vermittelte Nachweise von Forschungsleistungen erfüllen mindestens *auch* gesellschaftliche Erwartungen und dienen dazu, Spitzenforscherinnen zu identifizieren, Auszeichnungen zu vergeben, Professuren zu besetzen wie auch die Leistungsfähigkeit etwa einer Universität oder Disziplin zu beurteilen. Unmittelbare wie auch wiederum mittelbare Nachweise von Forschungsleistungen können zudem erkenntnisleitende Funktionen haben – gewissermaßen das Pendant zu didaktischen Funktionen von Prüfungen: Unmittelbare Forschungsergebnisse geben direkt Rückmeldung zum Forschungsprozess; Peer Reviews im Zuge der Vermittlung derselben sichern ein fachlich relevantes Feedback. Forschungsindikatoren wie *Impact Factor* und Zitationsindizes sind ähnlich wie Noten in der Prüfungspraxis Magneten der Aufmerksamkeit, reduzieren radikal die an sich gegebene Komplexität und nehmen großen Einfluss auf Forschende und ihre Karriere.

Sollte man Leistungen infolge des forschenden Lernens zwar erfassen, aber nicht benoten, um didaktische und erkenntnisleitende Funktionen des Prüfens zu befördern? Ist es angeraten, beim forschenden Lernen allein auf ein *Assessment for Learning* zu set-

7 Das heißt umgekehrt nicht, dass Prüfungen mit Rechtsfolgen immer benotet werden müssen: Je nach Prüfungsordnung bzw. rechtlicher Grundlage sind auch ein „bestanden – nicht-bestanden“ möglich.

zen? Oder sollte man beim forschenden Lernen Leistungen erfassen *und* benoten, um berechtigten gesellschaftlichen Erwartungen nachzukommen? Sind summativ benotete Prüfungen mit Rechtsfolgen gar eine passende Form des Leistungsnachweises – analog zu aggregierten Indikatoren der Forschungsleistung mit Karrierefolgen? Oder ist beides Ausdruck einer Fehlentwicklung, die zu hinterfragen und nicht noch zu verstärken ist? Ich denke nicht, dass eine hochschuldidaktische Theorie des Prüfens solche Fragen beantworten wird oder auch nur beantworten können soll. Sie muss aber Begriffe und Kategorien bereitstellen, um Antworten zu finden. Die Unterscheidung verschiedener Funktionen des Prüfens kann eine Hilfe sein, um sich im konkreten Fall klar zu werden: Welche Funktionen des Prüfens kann, will oder muss ich überhaupt erfüllen? Das wird in hohem Maße auch damit zu tun haben, wie das Lehrangebot zur Förderung forschenden Lernens in einem Studieneingang verankert ist (z. B. verpflichtend oder als Option), welches Gewicht es dort hat (z. B. in Form der Anzahl von Leistungspunkten) und welche Vorgaben (Modulhandbücher, Prüfungsordnungen) gegebenenfalls zu berücksichtigen sind.

4.2 Strukturen des Prüfens und mögliche Formen der Kohärenzstiftung

Setzt man sich mit Funktionen des Prüfens auseinander, wird einem schnell bewusst: Sie sind kaum trennscharf zu fassen. Was man ins Zentrum rückt, hängt davon ab, wer von welchem Blickwinkel aus mit welchen Erwartungen auf das Prüfen sieht. Versucht man, die verschiedenen Funktionen einmal abzustreifen und sich darauf zu konzentrieren, was man eigentlich macht, wenn man prüft, dann reduziert sich das Prüfen auf einen komplexen Akt mit zwei Teilen: Der Prüfende erfasst eine Leistung als Prozess oder Ergebnis und bewertet diese Leistung anhand bestimmter Maßstäbe. Allerdings – das muss man einschränkend dazu sagen – ist die Erfassung einer Leistung immer schon ein Akt der Interpretation; implizit ist die Bewertung im Prüfen geradezu eingewoben. Sie explizit zu machen und dann zu verbalisieren oder mit kompakten Symbolen – beispielsweise mit Ziffernnoten – zu versehen, lässt es aber gerechtfertigt erscheinen, das Bewerten vom Erfassen analytisch zu trennen.

Meine These an dieser Stelle ist, dass das die Prüfung konstituierende Erfassen und Bewerten ein Modus menschlichen Handelns ist, den wir beim Lernen und Forschen kennen, auch ohne in einer offiziellen Prüfungssituation zu sein. Wenn wir etwas lernen, brauchen wir Orientierung in mehrfacher Hinsicht – besonders dann, wenn wir versuchen, selbstorganisiert zu lernen, also frei von allen Zwängen und institutionellen Zumutungen. Wir bemerken und beobachten, was wir schrittweise erreichen, und setzen das ins Verhältnis zu dem, was wir eigentlich wollten. Wir beobachten unseren Fortschritt und machen uns ein Bild darüber, ob und wie wir uns verändern. Wir wollen wissen, was andere von dem halten, was wir erreichen, und wie wir damit im Vergleich zu anderen stehen. Wenn wir ein Forschungsvorhaben verfolgen, tun wir das – zumindest im Idealfall –, weil uns eine Frage umtreibt. Unseren Fortschritt messen wir daran, ob wir einer Antwort näherkommen. Neben unserer Neugier ist der bestehende For-

schungsstand eine wichtige Referenz, die wir heranziehen, um unser Tun zu lenken. Wir teilen Zwischen- und Endergebnisse anderen mit und suchen die Einschätzung der Fachgemeinschaft. Diese Beispiele zeigen vor allem informelle, wohl auch intuitive, Selbstbewertungsmechanismen, die aber immer auch der sozialen Einbettung bedürfen, also Fremdurteile direkt (z. B. Austausch mit anderen) oder indirekt (z. B. Vergleichsmaßstäbe über andere) mit einbeziehen.

Zu Beginn des Beitrags habe ich die Ausgangssituation für das Prüfen im Kontext forschenden Lernens skizziert und bin auf gängige Imperative der Gerechtigkeit eingegangen, die mit Qualität, Exzellenz und Diversität zu tun haben. Dahinter liegen verschiedene Bezugsnormen für das Prüfen, nämlich soziale, sachliche und individuelle. Diese Bezugsnormen, und das wollte ich hier zeigen, werden auch als Vergleichsmaßstäbe im offiziell ungeprüften Lern- und Forschungshandeln wirksam. Menschen machen in ihrem Handeln ständig und meist implizit und in unterschiedlicher Gewichtung die anderen, die Sache und sich selbst zum Maßstab und richten Entscheidungen an den Ergebnissen ihrer Vergleiche aus. Das ist aus didaktischer Sicht günstig, denn Prüfen und Geprüft-Werden sind so gesehen nichts, was man *zwingend* als wesensfremd, als künstlich oder aufgesetzt bezeichnen müsste. Es käme also beim Prüfen im Kontext des forschenden Lernens darauf an, an die „natürlichen“ Momente anzudocken, ein kohärentes Erleben von Lernen, Forschen und Prüfen anzustreben und eine „Entfremdung“ des Prüfens von unserem Handeln zu vermeiden.⁸ Ich möchte abschließend drei Impulse geben, um eine solche Kohärenz zu gewinnen und eine weitere entfremdende Entwicklung zu stoppen.

Der erste Impuls greift das *Prüfen analog zu Nachweisen von Forschungsleistungen* noch einmal auf. Die zugrundeliegende These lautet: Wer forschend lernt und sich auch bei Leistungsnachweisen in der Logik der Forschung weiß, wird dies als kohärent erleben und das Prüfen als Bestandteil des gesamten Prozesses nicht nur akzeptieren, sondern produktiv nutzen. Zu nutzen wäre also die Systematik zur Generierung von Prüfungsformaten, die sich an den analogen Prüfstrukturen des Forschungsprozesses orientiert. Die Herausforderung auf dem Weg des Prüfens analog zu Nachweisen von Forschungsleistungen ist: Einer authentischen Forschungssituation mit ihren Prüfmomenten möglichst nahekommen, ohne aber Entwicklungen in der Forschung zu kopieren, die ihrerseits die Forschenden von der Wissenschaft entfremden.

Die zweite Möglichkeit, von der ersten in gewisser Weise abhängig, bezeichne ich als *Prüfen in Mitbestimmung*. Die zugrundeliegende These ist: Wer beispielsweise mitbestimmen kann, wann seine Leistung im forschenden Lernprozess prüfungsreif ist, wird Prüfungen weniger als aufgesetzt empfinden und eher als Möglichkeit erkennen,

8 Ich gehe an der Stelle nicht weiter auf die aktuell z. B. von Mau (2017) treffend beschriebenen Exzesse und damit verbundenen Folgen für die beständige Vermessung nahezu aller Bereiche und Aktivitäten in unserer Gesellschaft infolge der digitalen Transformation und der dadurch möglich gewordenen Algorithmisierung ein. Auch diese wird natürlich durch die menschliche Neigung erst ermöglicht, sich mit anderen zu vergleichen und sich selbst (im Vergleich zu anderen) zu optimieren.

die eigene Forschungskompetenz zu entwickeln. Orientiert man sich an den generischen Prüfungsformen, kann man festhalten: Hausarbeiten – also z. B. Forschungsberichte – sind bereits ein typischer Fall, bei dem Studierende innerhalb eines gesetzten Zeitrahmens selbst bestimmen, wann sie ihr Werk zur Bewertung freigeben. In ähnlicher Weise dürfte das bei allen Formen von Materialisierung aus dem Forschungsprozess machbar sein. Portfolio-Prüfungen etwa gehen in genau diese Richtung. Da sich Vorträge und Demonstrationen auch aufzeichnen lassen, bestünde prinzipiell selbst hier die Möglichkeit, dass Studierende so lange daran feilen, bis sie ihr Leistungsniveau damit zeigen können. Klausuren und Gespräche dagegen entziehen sich weitgehend dem Modus des selbstbestimmten Prüfens. Die Herausforderung auf dem Weg des selbstbestimmten Prüfens ist die nicht kalkulierbare Selbstdisziplin der Studierenden.

Ein dritter Impuls erscheint mir bedenkenenswert immer dann, wenn Prüfungsfunktionen und -formen bereits gesetzt und dergestalt sind, dass man sie weder mit der Logik des Forschens in Einklang bringen noch im Modus der Selbstbestimmung umsetzen kann: die *Entkoppelung des Lehrens vom Prüfen*. Vor über zehn Jahren haben Peter Treppe und Balthasar Eugster den unkonventionellen Vorschlag gemacht, summative Prüfungen bzw. Prüfungen mit Rechtsfolgen von externen Experten durchführen zu lassen (Treppe & Eugster, 2006, S. 165). Das würde die an der Universität übliche Personalunion von Lehrenden und Prüfenden aufheben – eine Form, wie wir sie bei Staatsexamina zumindest kennen. Dahinter steckt die These, dass es das forschende Lernen entlastet, wenn sich Lehrende darauf konzentrieren, Studierende in ihrem Forschungsprozess anzuregen, zu unterstützen, ausschließlich formativ prüfend Feedback zu geben, und die abschließende, gegebenenfalls benotete Prüfung auslagern. Auf dem Weg der Entkoppelung des Lehrens vom Prüfen steht man vor der großen Herausforderung, lange und feste Traditionen in Frage zu stellen wie auch die damit verbundenen (nicht gewollten) Nebeneffekte in den Griff zu bekommen.

5. Abschließendes Fazit

Prüfungen haben – auch an Universitäten – nicht selten den Charakter von Ritualen. In den verschiedensten Kulturen kann man beobachten, wie Prüfungsrituale tradiert werden, die Bestand haben, auch wenn Kritiker noch so laut werden (Füssel, 2012, S. 101). Die seit 1900 wirkende Dominanz der Psychologie für das Verständnis von Prüfungen an Universitäten ist ein gutes Beispiel dafür: Tief verwurzelt und weitgehend immun gegen Kritik steuern psychologische Diagnostik- und Test-Modelle mit leichten Abwandlungen über die Zeit das Denken und Handeln im Lehr- und Prüfungsalltag. Der hehre Wunsch nach Gerechtigkeit macht empfänglich für das Versprechen, Kompetenzen objektiv, reliabel und valide erfassen zu können, indem man nur sorgfältig genug Aufgaben kreiert, die eine sichtbare Leistung hervorrufen, aus der man eine unsichtbare Kompetenz erschließen kann (Treppe & Reusser, 2007, S. 8; Wilbers, 2013, S. 302). In Kombination mit vor allem sachlichen und sozialen Bezugsnormen klingt das fair, angemessen und dazu noch wissenschaftlich. Gleichzeitig wissen wir, dass genau diese

Anforderungen an den Nachweis für den Erwerb von Kompetenzen in der Regel viel zu hoch und für den Regelbetrieb unrealistisch sind (vgl. Wex, 2012). Spätestens beim forschenden Lernen wird deutlich, wie wenig die scheinbar so treffende Lösung standardisierter Prüfungen in der Lage ist, die besondere Beziehung zwischen Lernen und Forschen zu berücksichtigen, wie sie akademische Lehre fördern sollte.

Es fehlt – so habe ich zum Einstieg dieses Beitrags festgestellt – eine hochschuldidaktische Theorie des Prüfens, die Hochschullehrende darin leiten könnte, Formen und Verfahren akademischen Prüfens zu finden, die den vielfältigen Verknüpfungen von Forschung und Lehre – auch und besonders dem *forschenden Lernen* – angemessen sind. Auf dem Weg zur Entwicklung einer Theorie habe ich zwei Richtungen eingeschlagen: Zum einen habe ich versucht, die manifesten Erscheinungen des Prüfens zu modellieren – mit zwei Systematiken, die helfen könnten, Prüfungsformate theoriegeleitet zu kreieren. Zum anderen habe ich damit begonnen, latente Bedeutungen des Prüfens zu analysieren – exemplarisch anhand der Funktionen und Strukturen des Prüfens wiederum mit einem Fokus auf dem forschenden Lernen: darin wird besonders deutlich, wie einseitig wir das Prüfen meist behandeln. Wenn, wie eingangs behauptet, die Zukunft des Prüfens historisch betrachtet offen ist, dann, so meine ich, gilt es, diese Zukunft auch zu gestalten. Mit den vorliegenden Ausführungen wollte ich genau dazu beitragen und zwar im Sinne des Leitgedankens einer Bildung durch Wissenschaft, dem sich das Konzept des forschenden Lernens, das ich hier ins Zentrum gerückt habe, nach wie vor verpflichtet fühlt.

Literatur

- Alexander, P. A. (2018). Past as Prologue: Educational psychology's legacy and progeny. *Journal of Educational Psychology*, 110(2), 147–162.
- Bornmann, L., & Marx, W. (2014). How to Evaluate Individual Researchers Working in the Natural and Life Sciences Meaningfully? A proposal of methods based on percentiles of citations. *Scientometrics*, 98(1), 487–509.
- Bruner, J. S., Olver, R. R., & Greenfield, P. M. (1966). *Studies in Cognitive Growth*. Oxford: Wiley.
- BAK = Bundesassistentenkonferenz (1970/2009). *Forschendes Lernen – Wissenschaftliches Prüfen*. Bielefeld: Webler.
- Füssel, M. (2012). Prüfungs- und Graduierungsrituale in unterschiedlichen Kulturen. Selektion – Initiation – Distinktion. In N. Huber, A. Schelling & S. Hornbostel (Hrsg.), *Der Dokortitel zwischen Status und Qualifikation* (iFQ-Working Paper 12, S. 101–108). Berlin: iFQ.
- Gerick, J., Sommer, S., & Zimmermann, G. (Hrsg.) (2018). *Kompetent Prüfungen gestalten*. Münster: Waxmann.
- Huber, L., & Reinmann, G. (im Druck). *Wege der Bildung durch Wissenschaft. Vom forschungsnahen zum forschenden Lernen an Hochschulen*. Berlin: Springer VS.
- Huber, L. (2009). Warum Forschendes Lernen nötig und möglich ist. In L. Huber, J. Hellmer & F. Schneider (Hrsg.), *Forschendes Lernen im Studium. Aktuelle Konzepte und Erfahrungen* (S. 9–35). Bielefeld: Webler.

- Huber, L. (2013). Die weitere Entwicklung des Forschenden Lernens. Interessante Versuche – dringliche Aufgaben. In L. Huber, M. Kröger & H. Schellhowe (Hrsg.), *Forschendes Lernen als Profilmerkmal einer Universität. Beispiele aus der Universität Bremen* (S. 21–36). Bielefeld: Webler.
- Huber, L. (2014). Forschungsbasiertes, Forschungsorientiertes, Forschendes Lernen: Alles dasselbe? *Hochschulforschung*, 1+2, 22–29.
- Iller, C., & Wick, A. (2009). Prüfungen als Evaluation der Kompetenzentwicklung im Hochschulstudium. *Das Hochschulwesen*, 57(6), 195–201.
- Kaminski, A. (2011). Prüfungen um 1900: Zur Genese einer Subjektivierungsform. *Historische Anthropologie*, 19(3), 331–353.
- Klauer, H. J., & Leutner, D. (2014). *Lehren und Lernen. Einführung in die Instruktionspsychologie*. Weinheim: Beltz.
- Macke, G., Hanke, U., Viehmann-Schweizer, P., & Raether, W. (2016). *Kompetenzorientierte Hochschuldidaktik: Lehren – vortragen – prüfen – beraten*. Weinheim: Beltz.
- Mau, S. (2017). *Das metrische Wir. Über die Quantifizierung des Sozialen*. Frankfurt a. M.: Suhrkamp.
- Mayer, R. E. (2018). Educational Psychology's Past and Future Contributions to the Science of Learning, Science of Instruction, and Science of Assessment. *Journal of Educational Psychology*, 110(2), 174–179.
- Murtonen, M., Gruber, H., & Lethinen, E. (2017). The Return of Behaviourist Epistemology: A review of learning outcome studies. *Educational Research Review*, 22, 14–128.
- Nagel, J. (2017). Prüfungspraktiken an Gymnasien und Lateinschulen in der Frühen Neuzeit. *Zeitschrift für Pädagogik*, 63(3), 259–279.
- Reeves, T. C. (2006). How Do We Know They Are Learning? The importance of alignment in higher education. *International Journal of Learning Technology*, 2(4), 294–309.
- Reinmann, G. (2016a). Gestaltung akademischer Lehre: semantische Klärungen und theoretische Impulse zwischen Problem- und Forschungsorientierung. *Zeitschrift für Hochschulentwicklung*, 11(5), 225–244.
- Reinmann, G. (2016b). Gestaltung akademischer Lehre. Anforderungen an eine Hochschuldidaktik als Allgemeine Didaktik. *Jahrbuch Allgemeine Didaktik* 11(2016), 45–60.
- Reinmann, G. (2017). Prüfungen und Forschendes Lernen. In H. A. Mieg & J. Lehmann (Hrsg.), *Forschendes Lernen. Wie die Lehre in Universität und Fachhochschule erneuert werden kann* (S. 115–128). Frankfurt a. M.: Campus.
- Reinmann, G. (2018). Shift from Teaching to Learning und Constructive Alignment: Zwei hochschuldidaktische Prinzipien auf dem Prüfstand. *Impact Free* 14. Hamburg.
- Ricken, N., & Reh, S. (2017). Prüfungen – Systematische Perspektiven der Geschichte einer pädagogischen Praxis. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 63(3), 247–259.
- Schaper, N., & Hilkenmeier, F. (2013). *Umsetzungshilfen für kompetenzorientiertes Prüfen* (HRK-Zusatzgutachten). <https://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-03-Material/zusatzgutachten.pdf> [27.03.2019].
- Tremp, P., & Eugster, B. (2006). Universitäre Bildung und Prüfungssystem – Thesen zu Leistungsnachweisen in modularisierten Studiengängen. *Das Hochschulwesen*, 5, 163–165.
- Tremp, P., & Hildbrand, T. (2012). Forschungsorientiertes Studium – universitäre Lehre: Das „Zürcher Framework“ zur Verknüpfung von Lehre und Forschung. In T. Brinker & P. Tremp (Hrsg.), *Einführung in die Studiengangentwicklung* (S. 101–116). Bielefeld: Bertelsmann.
- Tremp, P. & Reusser, K. (2007). Leistungsbeurteilung und Leistungsnachweise in Hochschule und Lehrerbildung – Trends und Diskussionsfelder. *Beiträge zur Lehrerbildung*, 25(1), 5–13.
- von Hentig, H. (1970). Wissenschaftsdidaktik. In H. v. Hentig, L. Huber & P. Müller (Hrsg.), *Wissenschaftsdidaktik* (Neue Sammlung Sonderheft 5, S. 13–40). Göttingen: Vandenhoeck & Ruprecht.

- Wex, P. (03.10.2012). Das leere Versprechen der Kompetenzprüfung. *Faz.net* <http://www.faz.net/aktuell/feuilleton/forschung-und-lehre/selbsttaeuschung-deruniversitaeten-das-leere-versprechen-der-kompetenzenpruefung-11910676.html> [27.03.2019].
- Wilbers, K. (2013). Kompetenzmessung: Motor der Theorie- und Praxisentwicklung in der Berufsbildung? In S. Seufert & C. Metzger (Hrsg.), *Kompetenzentwicklung in unterschiedlichen Kulturen* (S. 298–321). Paderborn: Eusl.
- Wissenschaftsrat (2017). *Strategien für die Hochschullehre. Positionspapier*. <https://www.wissenschaftsrat.de/download/archiv/6190-17.pdf> [27.03.2019].
- Wolter, F., & Schiener, J. (2016). Kompetenzdiagnostik an der Hochschule. Methoden, Herausforderungen und aktuelle Entwicklungen. In D. Großmann & T. Wolbring (Hrsg.), *Evaluation von Studium und Lehre. Grundlagen, methodische Herausforderungen und Lösungsansätze* (S. 367–402). Wiesbaden: Springer VS.

Abstract: Despite the ubiquity of university tests and examinations, we do not currently have a genuinely didactic theory of assessment. This is particularly noticeable when assessing and evaluating performances resulting from research-based learning in higher education. The article describes the difficulties and contradictory requirements in these situations, discusses ways of assessment analogous to research and tries to give impulses for the development of a theoretical framework for academic assessment.

Keywords: University Teaching, Research Orientation, University Assessment, Forms of Assessment, Learning Through Research

Anschrift der Autorin

Prof. Dr. Gabi Reinmann, Universität Hamburg,
Hamburger Zentrum für Universitäres Lehren und Lernen (HUL),
Leitung | Professur für Lehren und Lernen an der Hochschule,
Schlüterstraße 51, 20146 Hamburg, Deutschland
E-Mail: gabi.reinmann@uni-hamburg.de