



Testen und Bewerten

journal für lehrerInnenbildung

j | l | b
no. 1
2022

journal für lehrerInnenbildung

j | b

journal für lehrerInnenbildung

jlb

Herausgeber*innen

Prof. Dr. Caroline Bühler, Pädagogische Hochschule Bern
Prof. Dr. Eveline Christof, Universität für Musik und darstellende Kunst Wien
Prof. Dr. Bernhard Hauser (geschäftsführender Herausgeber),
Pädagogische Hochschule des Kantons St. Gallen
Prof. Dr. Kathrin Krammer, Pädagogische Hochschule Luzern
Prof. Dr. Jennifer Paetsch, Otto-Friedrich-Universität Bamberg
Prof. Dr. Kerstin Rabenstein, Georg-August-Universität Göttingen
Prof. Dr. Ilse Schrittemser, Universität Wien
Prof. Dr. Claudia Schuchart, Bergische Universität Wuppertal
Mag. Dr. Andrea Seel, Kirchliche Pädagogische Hochschule Graz

Redaktion

Stefanie Thurnheer, B. A.

Wissenschaftlicher und fachlicher Beirat

Prof. Dr. Herbert Altrichter, Johannes-Kepler-Universität Linz
Prof. Dr. Erwin Beck, Pädagogische Hochschule des Kantons St. Gallen
Prof. Dr. Ingrid Gogolin, Universität Hamburg
Prof. Dr. Tina Hascher, Universität Bern
Prof. Dr. Barbara Koch-Priewe, Universität Dortmund
Prof. Dr. Michael Schratz, Universität Innsbruck
Prof. Dr. Ursula Streckenisen, Pädagogische Hochschule Bern
Prof. Dr. Ewald Terhart, Universität Münster
Prof. Dr. Johannes Mayr, Alpen-Adria-Universität Klagenfurt

journal für lehrerInnenbildung

jlb

22. Jahrgang (2022)
Heft 1

Testen und Bewerten

Korrespondenzadresse der Herausgeber*innenredaktion:
Stefanie Thurnheer, B. A.
E-Mail: jlb.redaktion@phsg.ch

Rezensionen:
Tilman Drope, Dr.
Georg-August-Universität Göttingen
E-Mail: tilman.drope@sowi.uni-goettingen.de

jlb. journal für lehrerInnenbildung erscheint vier Mal im Jahr.
Sämtliche Ausgaben der jlb sind unter der Domain jlb-journallehrerinnenbildung.net
im Open Access auch online zugänglich.

Die Bezugsbedingungen zu einem Print-Abonnement finden Sie in unserem Webshop:
www.klinkhardt.de/verlagsprogramm/zeitschriften/

Bestellungen und Abonnentenbetreuung:
Verlag Julius Klinkhardt
Ramsauer Weg 5
83670 Bad Heilbrunn, Deutschland
vertrieb@klinkhardt.de
Tel: +49 (0)8046-9304
Fax: +49 (0)8046-9306
oder nutzen Sie unseren webshop:
www.klinkhardt.de

Dieser Titel wurde in das Programm des Verlages
mittels eines Peer-Review-Verfahrens aufgenommen.
Für weitere Informationen siehe www.klinkhardt.de.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten
sind im Internet abrufbar über <http://dnb.d-nb.de>.

2022.i. © Verlag Julius Klinkhardt.

Satz, Redaktion und Gestaltung: Elske Körber, Dipl.-Päd., München.

Druck und Bindung: AZ Druck und Datentechnik, Kempten.
Printed in Germany 2022.

Gedruckt auf chlorfrei gebleichtem alterungsbeständigem Papier.



*Die Publikation (mit Ausnahme aller Fotos, Grafiken und Abbildungen) ist
veröffentlicht unter der Creative Commons-Lizenz:
CC BY-NC-SA 4.0 International
<https://creativecommons.org/licenses/by-nc-sa/4.0/>*

<https://doi.org/10.35468/jlb-01-2022>
ISSN 1681-7028 (Print-Ausgabe)
ISSN 2629-4982 (Online-Ausgabe)

INHALT

EDITORIAL 7

BEITRÄGE 13

01 14

Petra Herzmann und Anke Liegmann
Mündliche Prüfungen im Kontext
des Forschenden Lernens. Diskurs und Praxis

02 24

David Bisang, Kerstin Bäuerlein und Corinne Wyss
Analysekompetenz von Studierenden
mit einem Video-Portfolio beurteilen

03 34

Christoph Kruse
Die Beurteilung angehender
Lehrer*innen im Referendariat.
Relevanz, Kontext und erste Explorationen

04 46

Kerstin Rabenstein
Einblicke in die soziale Praxis
des Prüfens an Hochschulen. Innen(ein)sichten

05 56

Michaela Kaiser
Mythos Eignungsprüfung.
Kunst – Leistung – hochschulische Prüfungspraxis

66	06 <i>Annette Bauer-Klebl und Charlotte Nüesch</i> Prüfungsevaluationen und damit verbundene Feedbackgespräche: Wie sie zur Hochschulentwicklung beitragen können
77	STICHWORT
78	07 <i>Bernhard Hauser</i> Testing Drives Learning?
89	REZENSION
97	AGENDA
99	CALL FOR ABSTRACTS

EDITORIAL

Bernhard Hauser
Kerstin Rabenstein

Prüfungen – und im engeren Sinne – Testen sind im Alltag der ersten beiden Phasen der Lehrer*innenbildung allgegenwärtig: Einzelne Veranstaltungen, Module und Ausbildungsabschnitte werden mit Prüfungen abgeschlossen. Prüfungsergebnisse können Einfluss auf weitere (Aus-)Bildungswege haben. Zudem können je nach universitärem Fach und Standort Aufnahmeprüfungen eine Eingangsselektivität der Studierenden herstellen. Mitunter wird die Durchlässigkeit von Studiengängen indirekt – durch Hürden, die die ersten Prüfungen darstellen – eingeschränkt. Somit prägen Prüfungen auch den Alltag von Studium und Lehre in vielfältiger Weise. So hieß es auch vor 14 Jahren im Editorial von Heft 1/2008 des Journals für Lehrer*innenbildung zum Thema „Bewerten und Prüfen“: „Bewerten und Prüfen gehören zum unverzichtbaren Handwerk in der Ausbildung von Lehrpersonen“ (Jäger und Hauser, 2008, S. 4, mit Verweis auf Jäger, 2007).

Diskutiert wurde in dem damaligen Heft u. a. eine vermutete Ideenarmut der Prüfungspraxis in der Lehrer*innenbildung, die Frage qualitativvoller Multiple-Choice-Aufgaben in Prüfungen, berufsberichtigende Prüfungen in den USA und mögliche Folgen wie Kompetenzillusionen durch zu sanfte Beurteilung. Verbunden war mit dem Thema zudem, dass gerade auch „Prüfungsmodi im Lehramt“ (ebd., S. 5) dazu beitragen könnten, dass bloß träges Wissen reproduziert würde (in Anlehnung an Bertschy, 2007). Wie geprüft und getestet wird, ist nicht nur zentral für das alltägliche Erleben von Lehren und Lernen in der Lehrer*innenbildung, sondern äußerst relevant für die Diskussion der Qualität der Lehrer*innenbildung.

Prüfungen und Testen – bzw. ihre Resultate und Effekte – spielen in verschiedener Hinsicht für den Einzelnen eine zentrale Rolle. In Prüfungen übernehmen Dozierende – zwangsläufig – unter anderem Selektionsverantwortung: Sie stellen sicher, dass Studierende über relevante Kompetenzen, für deren Aufbau die Hochschule beauftragt ist, auch in ausreichendem Ausmaß verfügen. Dozierende stimmen untereinander Prüfungsmodalitäten ab, bereiten Prüfungen vor, verbringen Zeit mit ihrer Bewertung. Studierende bzw. Referendare vergleichen Prüfungserfahrungen bei unterschiedlichen Lehrenden, tauschen mitunter Prüfungsaufgaben, bereiten sich auf Prüfungen vor und erwarten Mitteilungen über ihre Prüfungsergebnisse. Gesprochen wird über Prüfungen – auch unter Dozierenden und Studierenden bzw. Referendaren – demnach bei vielen Gelegenheiten. Seltener sind Prü-

fungsinhalte, Prüfungsformate und die Effekte von Prüfungen jedoch Gegenstand kollegialer Reflexionen.

Prüfungen und Testen könnten demnach auch institutionell für Studiengangs- bzw. Hochschulentwicklung eine wichtige Rolle spielen, sie werden in dieser Hinsicht aber noch selten in Augenschein genommen. Einem Tabu gleich wird – über alle Disziplinen hinweg – über die Benotungspraxis selbst kaum gesprochen, dabei wird bezüglich der Notengebung – gerade in Bezug auf das Lehramtsstudium – eine inflationäre Praxis konstatiert (Müller-Benedic & Grözinger, 2017). Gegenüber der hohen Relevanz wird die (eigene) Praxis des Prüfens und Testens in der Lehrer*innenbildung vergleichsweise selten zum Gegenstand einer (fundierten) Diskussion gemacht.

Das vorliegende Heft will zur Diskussion anregen, indem es Einblicke in aktuelle Diskurse und Befunde zum Prüfen und Testen an Hochschulen im Allgemeinen und in lehramtsbezogenen Studiengängen bzw. pädagogischen Hochschulen und im Referendariat im Besonderen gewährt.

Unter die Lupe genommen werden im Folgenden zum einen spezifische Elemente der Lehrer*innenbildung, wie Prüfungen zu schulpraktischen Anteilen der Ausbildung in der ersten Phase (Herzmann & Liegmann; Bisang, Bäuerlein & Wyss) bzw. der zweiten Phase (Kruse). Zum zweiten ermöglichen Beschreibungen von Erfahrungen im hochschulischen Alltag des Prüfens Reflexionen dieser Praxis im Allgemeinen (Rabenstein) bzw. mit Bezug auf Eignungsprüfungen an Hochschulen der Künste im Besonderen (Kaiser). Des Weiteren wird die Rolle von Prüfungen und Testen für das Lehren und Lernen in Bezug auf Prüfungssysteme als Stellgrößen der Studiengangsentwicklung ebenfalls im Spezifischen (Bauer-Klebl & Nüesch) und im Allgemeinen (Hauser) aufgezeigt.

Petra Herzmann und *Anke B. Liegmann* befassen sich mit mündlichen Prüfungen zum Abschluss von sogenannten Studienprojekten, die im Rahmen schulpraktischer Anteile in der universitären Phase der Lehramtsausbildung in Deutschland durchgeführt werden. Studienprojekte sind dabei im Sinne einer von den Studierenden im Praxissemester an der Schule bzw. im Unterricht durchgeführte Lehrforschung zu verstehen. Mittels einer Analyse transkribierter mündlicher Prüfungsgespräche fragen die Autorinnen angesichts der Erwartungen an ein selbstständig zu vollziehendes Lehrforschungsprojekt nach der Herstellung von Prüfungsförmigkeit. Die Ergebnisse zeigen, wie sich Studierende

in den einerseits als (forschungs-)kompetent und andererseits als in Bezug auf die Praxis des Unterrichts lernende Noviz*innen darstellen. Die Ergebnisse werden für eine hochschuldidaktische Reflexion des Prüfungsformats genutzt.

David Bisang, Kerstin Bäuerlein und Corinne Wyss überprüfen die inhaltliche und ökologische Validität eines benoteten Video-Portfolios als – je nach Lesart – Ersatz (Verlust) beziehungsweise Weiterentwicklung der über Jahrzehnte eingesetzten Lehrprobe oder Prüfungslektion. Insgesamt findet sich eine hohe inhaltliche und eine mittlere ökologische Validität, was derartige Evaluationen zu einem wirksamen Instrument der Weiterentwicklung von Prüfungen macht.

Christoph Kruse untersucht eine noch kaum diskutierte Prüfungspraxis, die schriftliche Beurteilung angehender Lehrpersonen im Referendariat. Er stellt Ansatz und explorativ an ersten Daten den methodischen Zugang seiner Grounded Theory Studie zur Prüfungspraxis mit schriftlichen Gutachten von Fach- und Schulleistungen im Referendariat dar. Dabei arbeitet er unter anderem am Beispiel der Klassenführung heraus, dass zum einen – unter Bezugnahme auf den Kompetenzbegriff – eine grundsätzliche Erlernbarkeit der wichtigsten Berufskompetenzen postuliert werde, dass aber im Gegensatz dazu in starkem Ausmaß Persönlichkeitsaspekte als Erklärung von Berufserfolg in den Gutachten herangezogen würden, deren Erlernbarkeit – zumindest in der kurzen Zeit – fraglich ist.

Kerstin Rabenstein trägt Befunde ethnographischer Studien zum Prüfungsalltag an Hochschulen zusammen. Sie zeigt auf Basis von Einblicken in den hochschulischen Alltag, wie die hohe Prüfungsdichte in den modularisierten Studiengängen das Lehren und Lernen prägen. Dabei wird auch deutlich, dass Prüfungen zwar einerseits vermehrte Aufmerksamkeit in der Programmatik der Hochschulen gewidmet wird, doch andererseits Prüfungen als Belastung wahrgenommen werden und ein Rückzug von Hochschullehrenden und Studierenden aus Prüfungen zu beobachten ist. Thematisiert werden auf diese Weise Spannungen und Ambivalenzen, die den von Prüfungen durchzogenen Lehr- und Studieralltag – nicht nur der lehramtsbezogenen Studiengänge – betreffen.

Michaela Kaiser versteht ihren Beitrag zu den Routinen und ungestellten Fragen zur kunstpraktischen Eignungsprüfung im Studium des künstlerischen Lehramts in Deutschland als Problemaufriss. Bestehend aus Mappenprüfung, praktischer Prüfung und Gespräch bleibt

es doch tendenziell unklar, wie die Feststellung der künstlerischen Eignung zustande kommt oder kommen soll. Auf Basis eines Fachgesprächs und auf der Grundlage der Sichtung exemplarisch ausgewählter Studienordnungen gibt sie detaillierte Einblicke in dieses Prüfungsformat und die mit ihm einhergehenden offenen Fragen.

Annette Bauer-Klebl und *Charlotte Nüesch* stellen auf der Grundlage zehnjähriger Erfahrung mit der Evaluation summativer Prüfungen dar, wie mit der Veränderung des Prüfungssystems eines Studiengangs auf die Verbesserung der Lehrqualität Einfluss genommen werden kann. Sie legen dar, wie nicht nur die Lehre, sondern auch die Prüfungen institutionell evaluiert werden und somit Prüfungen als Instrument zur Kompetenzentwicklung und Evaluationen als Innovationsförderung zu verstehen sind.

Bernhard Hauser setzt sich mit den Effekten des Testens für das Lernen auseinander. Er gibt einen Überblick über Befunde zur Verbesserung von Lernen durch Testen. Es werden Wirkungen von Vergleichstests, von Teaching-to-the-Test und formativen Tests beschrieben. Dabei wird unter anderem konstatiert, dass Teaching-to-the-Test nur dann ein Problem ist, wenn die Tests schlecht sind.

Literatur

- Bertschy, B. (2007). Pädagogisches Wissen problemorientiert prüfen. *Beiträge zur Lehrerbildung*, (1), 37-50.
- Jäger, R. S. (2007). *Beobachten, bewerten, fördern! Lehrbuch für die Aus-, Fort- und Weiterbildung*. Landau: Verlag Empirische Pädagogik.
- Jäger, R. S. & Hauser, B. (2008). Editorial. Bewerten und Prüfen. *journal für lehrerInnenbildung*, (1), 4-6.
- Müller-Benedic, V. & Grözinger, G. (Hrsg.). (2017). *Noten an Deutschlands Hochschulen. Analysen zur Vergleichbarkeit von Examensnoten 1960 bis 2013*. Wiesbaden: Springer VS.



Bernhard Hauser, Prof. Dr. phil.
an der Pädagogischen Hochschule St. Gallen.
Arbeitsschwerpunkte:
Lehren und Lernen,
Lernwirksamkeit von Spiel,
Bildungsforschung bei 3- bis 10-Jährigen

bernhard.hauser@phsg.ch



Kerstin Rabenstein, Dr., Professorin
am Institut für Erziehungswissenschaft,
Georg-August-Universität Göttingen.
Arbeitsschwerpunkte:
Transformation pädagogischer Ordnungen,
qualitative Schul- und Unterrichtsforschung

kerstin.rabenstein@sowi.uni-goettingen.de

01

Petra Herzmann und Anke Liegmann
Mündliche Prüfungen im Kontext
des Forschenden Lernens. Diskurs und Praxis

02

David Bisang, Kerstin Bäuerlein und Corinne Wyss
Analysekompetenz von Studierenden
mit einem Video-Portfolio beurteilen

03

Christoph Kruse
Die Beurteilung angehender
Lehrer*innen im Referendariat.
Relevanz, Kontext und erste Explorationen

04

Kerstin Rabenstein
Einblicke in die soziale Praxis
des Prüfens an Hochschulen. Innen(ein)sichten

05

Michaela Kaiser
Mythos Eignungsprüfung.
Kunst – Leistung – hochschulische Prüfungspraxis

06

Annette Bauer-Klebl und Charlotte Nüesch
Prüfungsevaluationen und
damit verbundene Feedbackgespräche:
Wie sie zur Hochschulentwicklung
beitragen können

01

*Petra Herzmann und
Anke B. Liegmann*

Mündliche Prüfungen im Kontext
des Forschenden Lernens.
Diskurs und Praxis

Das Forschende Lernen ist nicht zuletzt durch die Einführung verlängerter Praxisphasen – wie dem sogenannten Praxissemester – erneut zu einem in Bildungspolitik und Erziehungswissenschaft intensiv diskutierten *hochschuldidaktischen Ansatz* der universitären Lehrerbildung avanciert. Im Rahmen der folgenden Ausführungen werden wir uns spezifisch mit mündlichen Prüfungen beschäftigen, die den Abschluss einer zweisemestrigen, forschungsbezogenen Auseinandersetzung der Student*innen mit Schule und Unterricht, während und nach dem Praxissemester, darstellen.¹ Ausgehend von programmatischen Erwartungen an das Forschende Lernen legen wir Befunde zur Prüfungspraxis zusammenfassend dar, die wir im Rahmen unseres Forschungsprojekts „Studienprojekte im Praxissemester“ (StiPS) gewonnen haben (vgl. Herzmann & Liegmann, 2020a). Angesichts unserer Befunde erscheint es uns abschließend relevant, auf hochschuldidaktische Möglichkeiten zur Prüfungsgestaltung hinzuweisen.

(Reform-)Erwartungen an das Forschende Lernen

Mit dem Forschenden Lernen im fortgeschrittenen Verlauf des Lehramtsstudiums ist im Praxissemester in Nordrhein-Westfalen verbunden, dass die Studierenden (mindestens) ein eigenes, in seinem Umfang begrenztes Forschungsvorhaben planen, an ihrer Praktikumsschule umsetzen und schriftlich dokumentieren. Dieses *Studienprojekt² bildet den zentralen Gegenstand der das Praxissemester abschließenden Prüfung, die häufig in Form einer mündlichen Prüfung durchgeführt wird. Das Studienprojekt wird von der Universität verantwortet; die Prüfung findet demzufolge an der Universität statt.* An das Forschende Lernen im Kontext der Lehrer*innenbildung werden insgesamt vielfältige und weitreichende Erwartungen geknüpft (vgl. z. B. Fichten, 2017; Herzmann & Liegmann, 2020b). Diese werden wir im Folgenden skizzenhaft ausweisen, um den Reformanspruch als normativen Bezugsrahmen auszuweisen, innerhalb dessen sich die

- 1 Studienprojekte in NRW werden in der Regel in einem Semester geplant und im darauffolgenden Semester umgesetzt. Die Prüfung steht am Ende dieses zweisemestrigen Zyklus.
- 2 An vielen Standorten in NRW werden zwei Studienprojekte durchgeführt, an einigen nur eines. Wenn zwei Studienprojekte durchgeführt werden, sind sie in zwei unterschiedlichen Fächern resp. Bildungswissenschaften verortet. Gegenstand der Prüfung ist jeweils nur ein Studienprojekt.

Akteure im Praxissemester – wie unsere Untersuchung zeigt – in spezifischer Weise bewegen.

Forschende Selbsttätigkeit

Im Diskurs um das Forschende Lernen ist umstritten, wie eng die studentische Forschung sich an den Maßgaben empirischer Forschung zu orientieren habe. Unstrittig aber ist, dass der Einsatz sozialwissenschaftlicher Methoden eine hohe Plausibilität gewinnt, denn sie ermöglichen – je nach Methode zu unterschiedlichen Fragestellungen – eine an wissenschaftlichen Geltungsansprüchen orientierte Auseinandersetzung mit der schulischen bzw. beruflichen Praxis. Als Erwartung an Studierende ist damit zum Ausdruck gebracht, dass forschungsmethodisches Wissen in der Bearbeitung des Studienprojekts zur Anwendung kommen muss, womit ein *vergleichsweise hohes Maß an forschender Selbständigkeit* auf Seiten der Studierenden vorausgesetzt ist. Dass die Studierenden dabei einer eigenen Frage- oder Problemstellung nachgehen, verweist auf ein programmatisches Kernstück Forschenden Lernens. So betont Huber (2009) die Autonomie der Studierenden beim Forschenden Lernen, die sich auch in einer selbstständigen Wahl des Themas niederschlagen sollte. Auch Katenbrink und Wischer (2014) verweisen darauf, dass es „eine Gelingensbedingung für Forschendes Lernen [ist], das Studierende gerade beim Finden einer Forschungsfrage die Möglichkeit haben müssen, ihre Interessen einzubringen“ (ebd., S. 121). Selbst wenn in der universitären Begleitung der Studienprojekte davon abgewichen wird und bspw. Forschungsgegenstände oder Forschungsmethoden vorgegeben bzw. eingeschränkt werden, wird programmatisch ein *vergleichsweise hohes Interesse der Studierenden* im Hinblick auf das Studienprojekt vorausgesetzt.

Transfer in die berufliche Praxis

Mit dem Anschluss an die Programmatik Forschenden Lernens wird neben diesen Forschungsleistungen im engeren Sinne auch der Transfer der mit der Forschung erzielten Ergebnisse in die berufliche Handlungspraxis in Schule und Unterricht gefordert (vgl. Fichten, 2017). Die empirisch informierte Reflexion des Handlungsfeldes Schule verweist auf die dem Forschenden Lernen zugrundeliegende Idee einer Theorie-Praxis-Verknüpfung. Professionstheoretisch zielt das Forschende Lernen in diesem Sinne auf die Entwicklung eines wissenschaftlich-

reflexiven Habitus angehender Lehrer*innen (Helsper, 2001), womit angedeutet ist, dass professionelles Handeln neben dem praktischen Können einer theoriebezogenen Reflexion des schulischen bzw. unterrichtlichen Handlungsfeldes und des eigenen (Lehrer*innen-)Handelns bedarf. Programmatisch rückt mit der Erwartung an eine Theorie-Praxis-Relationierung demzufolge nicht nur der Gegenstand der Forschung, sondern auch die forschende Person und ihre Reflexionsfähigkeit in den Fokus des Geschehens.

Dass diese (Reform-)Erwartungen an das hochschuldidaktische Format *Forschendes Lernen* nicht ohne Folgen für mündliche Prüfungen sind, stellen wir nun anhand von Befunden zur Praxis mündlicher Prüfungen dar.

Befunde zur Praxis mündlicher Prüfungen

Unsere adressierungsanalytischen Rekonstruktionen von 33 audio-graphierten Prüfungsgesprächen (zum methodischen Vorgehen vgl. ausführlich Herzmann & Liegmann, 2020a) weisen als zentrale Anforderung für die Akteure aus, sich in der Ambivalenz zwischen der Demonstration und Gewährung studentischer Expertise im Hinblick auf das Studienprojekt einerseits und deren prüfungsförmiger Begrenzung andererseits zu positionieren. Dass es sich um eine konventionelle Prüfungssituation im Kontext der universitären Lehrer*innenbildung handelt, der sich die Beteiligten mindestens bis zu einem Grad zu fügen haben, jedenfalls dann, wenn sie die Fortsetzung der Situation als Prüfung nicht gefährden wollen, steht dabei außer Frage. Interessiert hat uns vielmehr, wie Prüfungsförmigkeit im konkreten Vollzug von den Akteuren hergestellt wird, inwiefern sich die Beteiligten auf Prüfungsnormen beziehen bzw. diese situativ verhandeln und ob sich dadurch Geltungsansprüche in und an Prüfungen verändern (zum methodischen Vorgehen vgl. Rose & Ricken, 2018). Wir werden dies im Folgenden an zwei, für die hochschuldidaktische Gestaltung mündlicher Prüfungssituationen zentralen Herausforderungen kenntlich machen.

Zum Expertise-Status der Studierenden

Vor allem in der Eröffnung der Prüfungssituation zeigt sich der Versuch der Prüfer*innen, die Prüfung als Gespräch unter Expert*innen

zu rahmen und einen (fach-)wissenschaftlichen Diskurs zu initiieren. Die Prüfungssituation wird zu Beginn offen angelegt, indem den Studierenden die Gelegenheit eingeräumt wird, ihr Studienprojekt in Form eines Kurzvortrags darzustellen. Dabei wird bereits zu Beginn auf den Prüfungsteil verwiesen, der sich an den einleitenden Vortrag der Studierenden anschließt und der als das dann folgende „wirkliche Gespräch“³ bezeichnet wird. In der Anmoderation angesprochen wird auch das von den Studierenden bis zur Prüfung bereits gezeigte Engagement in der Durchführung des Studienprojekts. Die Prüfung wird als dessen „krönender Abschluss“ und „Höhepunkt“ benannt, was den Eindruck erweckt, dass das Prekäre einer Prüfungssituation, nämlich an ihr zu scheitern, kommunikativ abgeschwächt werden muss. Die Betonung des studentischen Engagements sowie die zumindest unterstellte Identifikation mit dem Forschungsvorhaben, indem bspw. das Interesse der Studierenden an der Fragestellung des Studienprojekts angesprochen wird, kann als Ausdruck der Anerkennung forschender Selbsttätigkeit (s. o.) interpretiert werden.

Die Studierenden demonstrieren im Kurzvortrag ihre forschungs- und fachbezogene Expertise (vgl. Meer, 2019). Entlang von vorab mit den Dozierenden abgestimmten (Darstellungs-)Kriterien, die in der Prüfungssituation nicht mehr benannt werden (müssen), zeigen die Geprüften ihr Wissen. Zugleich wird die Prüfungsförmigkeit der Situation von den Studierenden durch das Befolgen dieser Kriterien sowie durch rituelle Unterwerfungsgesten hergestellt („... ich neige dazu, zu schnell zu sprechen, wenn ich aufgeregt bin, was der Fall ist“).

Im weiteren Verlauf der Prüfung zeigt sich, dass auch die Prüfer*innen in einen Prüfungsmodus wechseln, der z. B. durch die Steuerung des weiteren Prüfungsverlaufs in Form des beanspruchten Fragerechts deutlich wird. Dabei werden die Studierenden weiterhin als Expert*innen adressiert, indem fachliche Fragen verhandelt werden oder die Expertise der Studierenden als Praktiker*innen (etwa auf die Frage der Gestaltung von Lernprozessen oder der Unterrichtsplanung) verschoben wird. Diese Expertise wird nur eingeschränkt, wenn im Vortrag deutliche Fehlkonzepte der Studierenden zum Vorschein kommen, die dann korrigierend überprüft werden.

3 Diese und die folgenden Äusserungen stammen aus dem Datenmaterial des Forschungsprojekts StiPS (Studienprojekte im Praxissemester), an dem neben den beiden Autorinnen auch Michaela Artmann und Marie Berendonk beteiligt waren.

Zum reflexiven Habitus der Studierenden

Programmatisch besteht die Erwartung, sich im Kontext des Forschenden Lernens als angehende Lehrer*innen reflektiert hinsichtlich des eigenen oder des bei den Lehrkräften der Praktikumsschule beobachteten Handelns zu zeigen. So stellt der Punkt *Reflexion* auch einen Teil der vorab festgelegten Gliederung in der Präsentation der Studienprojekte dar. Entlang der unterschiedlichen Einbettung der Studienprojekte in Unterrichtsaktivitäten (eigener Unterricht vs. Prozesse und Produkte des Unterrichts anderer Lehrkräfte) können wir zwei unterschiedliche Modi der Demonstration dieses reflexiven Anspruchs rekonstruieren: (1) Anfänger*innen-Sein und (2) Praktisch-Lernende.⁴

Studierende, die im Rahmen des Studienprojektes den *eigenen Unterricht* beforscht haben, positionieren sich als *Anfänger*innen im Feld des Lehrer*innenseins* (1) und stellen dabei die Lehrer*in-Schüler*in-Beziehung in den Vordergrund ihrer Ausführungen. Dabei nehmen sie entweder Bezug auf die von ihnen beiläufig wahrgenommene Schüler*innenmeinung oder sie holen sich Feedback der Schüler*innen zu dem von ihnen durchgeführten Unterricht ein. Die Rückmeldungen der Schüler*innen sind vor allem auf allgemeine Fragen der Lernatmosphäre bezogen (haben Spaß, sind motiviert, haben keine Angst). Die darin zum Ausdruck kommende Frage der Anerkennung durch die Schüler*innen kann demzufolge als eine elementare selbst gestellte Erwartung der Studierenden verstanden werden. Die eingenommene Anfänger*innenposition wird auch darin deutlich, dass Studierende eigene Fehleinschätzungen und -entscheidungen markieren (zu wenig Zeit eingeplant, Lehrmaterial falsch gewählt, Planung zu aufwändig). Wissenschaftliche Wissensbestände werden insofern als Referenz herangezogen, als theoretische Modelle (z. B. der Kommunikation) auf ihre Plausibilität hin (z. B. anhand der erforschten Kommunikationsstrukturen) geprüft werden. Ist hingegen nicht der eigene, sondern *fremder Unterricht* (bzw. dessen Prozesse und Produkte) Gegenstand des Studienprojektes, wird dieser zum Anlass genommen, Handlungsmaximen für die eigene (zukünftige) Unterrichtspraxis abzuleiten bzw. zu entwickeln. Die Studierenden positionieren sich damit als *am (praktischen) Beispiel Lernende* (2), die sich in die Rolle einer Lehrkraft versetzen bzw. allgemeine Handlungserwartungen von Lehrkräften antizipieren.

4 Diese Befunde sind in der Publikation Herzmann & Liegmann (2020a) noch nicht berücksichtigt.

Die Prüfer*innen wiederum re-adressieren in ihren Nachfragen Studierende als Lernende, die fachdidaktische Wissensbestände zu den Befunden ihres Studienprojektes (und der erlebten Praxis) in Beziehung setzen (können). Dies zeigt sich z. B. in Fragen nach fachdidaktischen Prinzipien und daraus abzuleitenden Begründungen für lehrerseitige Entscheidungen, aber auch in Rückfragen zu diagnostischen Einschätzungen und daraus resultierenden Empfehlungen bzw. Alternativen für das Lehrkräftehandeln.

Entgegen der weitreichenden Erwartungen an die Entwicklung eines wissenschaftlich-reflexiven Habitus zeigt sich in den von uns untersuchten Prüfungen, dass die Anforderung der Reflexion seitens der Prüfer*innen auf eine konkrete Relationierung von beobachteter Praxis und fachdidaktischem Wissen begrenzt ist. Ein Nachdenken über sich selbst in der Rolle der Lehrkraft im Sinne einer allgemeinen, habituellen Reflexionsfähigkeit – die etwa Fragen nach Anerkennungsverhältnissen der Studierenden gegenüber den Schüler*innen aufwirft – wird von den Studierenden weder demonstriert noch von den Dozierenden eingefordert.

Hochschuldidaktische Überlegungen

Angesichts der aufgezeigten Forschungsbefunde mündlicher Prüfungen im Kontext Forschenden Lernens werden wir nun hochschuldidaktisch konkretisieren, wie Prüfungssituationen gestaltet werden können. Wir stellen dazu Überlegungen anhand von zwei Fragestellungen an:

Was wird geprüft?

Mit den weitreichenden Erwartungen, die an das Forschende Lernen gestellt werden, ist die Bandbreite der Themen, die in mündlichen Prüfungen behandelt werden kann, als umfassend zu bezeichnen: Forschungsmethodische Fragen und fachliche bzw. fachdidaktische Perspektiven, die für das Studienprojekt relevant sind, bieten Möglichkeiten für Nachfragen und weiterführende Überlegungen. Dabei sind nicht nur die Forschungsergebnisse von Interesse, sondern auch der Forschungsprozess und die von den Studierenden getroffenen Entscheidungen, die im Sinne eines *doing research* in Prüfungen rückblickend zur Rede stehen können. Vorab festgelegte Gliederungspunkte für den studentischen Vortrag sowie Kriterien(-kataloge) der Prüfen-

den schaffen Transparenz über die formale Logik und Erwartungshorizonte mündlicher Prüfungen.

Der im professionstheoretischen Diskurs um das Forschenden Lernen so prominent vertretene Anspruch an reflexive Inszenierungspraktiken der Studierenden bzw. die Demonstration eines wissenschaftlich-reflexiven Habitus sind u. E. weniger auf selbstbezügliche Verhaltenserwartungen an Studierende im Hinblick auf Engagement und Identifikation zu beziehen. Vielmehr sind die skizzierten gegenstands- und forschungsmethodischen Begründungen im Forschungsprozess zu prüfen, bzw. es steht zur Disposition, inwiefern sich ausgehend von den erzielten Befunden Konsequenzen für z. B. die Diagnose von Lernständen, der Planung von Unterricht und Analyse von Verantwortlichkeiten in der (Einzel-)Schule formulieren lassen. Damit wären weniger die habituellen Reflexionen der Studierenden Gegenstand mündlicher Prüfungen. Vielmehr läge der Prüfung eine sachbezogene Kommunikation zugrunde, die Theorien als notwendige, aber keineswegs hinreichende Analysefoki von Praxis bestimmt. Die Bezugnahme von Wissen auf Praxis (und umgekehrt) erfolgte in diesem anspruchsvollen Sinne reflexiv (vgl. Herzmann, 2021).

Wie wird geprüft?

Für studentische Forschungsprojekte angemessen erscheint uns die Adressierung der zu Prüfenden als Expert*innen für ihr Studienprojekt, die sich beispielweise im anfänglichen Vortrag und in relativ hohen Redeanteilen der Studierenden auch im zweiten Teil der Prüfung zeigt. Dass es dialogische Elemente in der Prüfung gibt und diese von den Prüfenden wertschätzend anmoderiert werden, ist unseres Erachtens zu Recht dem Studienprojekt als einem vergleichsweise neuen und im beschriebenen Sinne besonderen Format im Kontext der universitären Lehrer*innenbildung geschuldet. Damit verbunden ist ebenfalls die i. d. R. längere Involviertheit der Prüfenden in die Beratung und Begleitung des Studienprojekts. Die mündliche Prüfung nimmt auf eine gemeinsame Vorbereitungspraxis der Beteiligten Bezug, sodass für mündliche Prüfungen im Kontext Forschenden Lernens – in einer gewissen Struktur analogie zu Disputationen im Rahmen von Promotionsverfahren – anzunehmen ist, dass nicht nur die studentische Forschungs-, sondern auch die Betreuungsleistung der Dozierenden zur Disposition steht (vgl. Meer, 2019).

Fazit

Dennoch bleibt eine mündliche Prüfung eine institutionell präfigurierte Situation studentischer Leistungsüberprüfung, die im Hinblick auf die Angemessenheit der Prüfungskriterien für studentisches Forschen bzw. Forschendes Lernen – wie wir zeigen konnten – durchaus gestaltbar ist. Als Prüfungskommunikation ist die mündliche Prüfung aber nicht als authentischer Abschluss einer ebensowenig authentischen Forschungserfahrung (miss-)zu verstehen – wie dies im Diskurs Forschenden Lernens gelegentlich formuliert wird (vgl. Reinmann, 2019). Prüfen bleibt eben der „Inbegriff der in sich spannungsreichen Logik des Pädagogischen“ (Ricken & Reh, 2017, S. 248). In anderen Worten: Eine Prüfung bleibt eine Prüfung, unabhängig davon, wie Prüfen gestaltet wird.

Literatur

- Fichten, W. (2017). Forschendes Lernen in der Lehrerbildung. In R. Schüssler, A. Schöning, V. Schwier, S. Schicht, J. Gold & U. Weyland (Hrsg.), *Forschendes Lernen im Praxissemester. Zugänge, Konzepte, Erfahrungen* (S. 30-38). Bad Heilbrunn: Klinkhardt.
- Helsper, W. (2001). Praxis und Reflexion. Die Notwendigkeit einer „doppelten Professionalisierung“ des Lehrers. *Journal für Lehrerinnen- und Lehrerbildung*, 3(1), 7-15.
- Herzmann, P. (2021). Zum Theorie- und Empirieanspruch (in) der Auseinandersetzung mit Praxis. Hochschuldidaktische Überlegungen zur Unhintergebarkeit von Nützlichkeitsforderungen an die universitäre Lehrer*innenbildung. *PFLB Zeitschrift für Schul- und Professionsentwicklung*, 3(5), 50-58.
- Herzmann, P. & Liegmann, A. B. (2020a). Mündliche Prüfungen im Kontext des Forschenden Lernens. (Re-)Adressierungen als Inszenierung studentischer Expertise. *Zeitschrift für Pädagogik*, 66(5), 727-745.
- Herzmann, P. & Liegmann, A. B. (2020b). Wie reflexionsförderlich sind universitäre Praxisphasen? Kritische Anmerkungen zu einem Professionalisierungsversprechen aus professionstheoretischer und empirischer Perspektive. In K. Rheinländer & D. Scholl (Hrsg.), *Verlängerte Praxisphasen in der Lehrer*innenbildung: Konzeptionelle und empirische Aspekte der Relationierung von Theorie und Praxis* (S. 74-88). Bad Heilbrunn: Klinkhardt.
- 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: Univ.-Verl.
- Katenbrink, N. & Wischer, B. (2014). Konzepte forschenden Lernens in der Osnabrücker Lehrerbildung. Versuch einer Einordnung und Reflexion. In N. Katenbrink, B. Wischer & Y. Nakamura (Hrsg.), *Forschendes Lernen in der Osnabrücker Lehrerbildung. Konzepte und Erfahrungen* (S. 109-131). Münster: MV-Wissenschaft.

- Meer, D. (2019). „Der Prüfer ist immer noch nicht der König“. Nachdenken über mündliche Hochschulprüfungen der Gegenwart. In T. Tyagunova (Hrsg.), *Studentische Praxis und universitäre Interaktionskultur. Perspektiven einer praxeologischen Bildungsforschung* (S. 173-199). Wiesbaden: Springer VS.
- Reinmann, G. (2019). Forschendes Lernen prüfen. Hochschuldidaktische Gedanken zu einer Theorie des Prüfens. *Zeitschrift für Pädagogik*, 65(4), 608-626.
- Ricken, N. & Reh, S. (2017). Prüfungen—Systematische Perspektiven der Geschichte einer pädagogischen Praxis. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 63(3), 247-258.

Petra Herzmann, Dr., Professorin
am Department Erziehungs- und Sozialwissenschaften,
Universität zu Köln.
Arbeitsschwerpunkte:
Professionalisierung und Professionalität
von Lehrpersonen,
Unterricht,
Rekonstruktive Forschungsmethoden



officeherzmann@uni-koeln.de

Anke B. Liegmann, Dr., wiss. Mitarbeiterin
in der Arbeitsgruppe Bildungsforschung,
Fakultät für Bildungswissenschaften,
Universität Duisburg-Essen.
Arbeitsschwerpunkte:
Professionalisierung von Lehrkräften
im Kontext von Diversität,
Praxisphasen in der Lehrkräftebildung



anke.liegmann@uni-due.de

02

*David Bisang,
Kerstin Bäuerlein
und Corinne Wyss*

Analysekompetenz
von Studierenden mit einem
Video-Portfolio beurteilen

Das Video-Portfolio der PH FHNW

Lehrprobe, Prüfungslektion oder schulpraktische Prüfung – an vielen Pädagogischen Hochschulen ist es üblich, dass die Studierenden am Ende ihrer schulpraktischen Ausbildung Unterrichtslektionen halten, die summativ beurteilt werden. Oft finden diese Prüfungen vor Ort in einer Schule mit Prüfer*innen statt (z. B. Esslinger-Hinz, 2016). An der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz (PH FHNW) gibt es seit 2014 eine alternative Beurteilungsform: das sogenannte Video-Portfolio (Bäuerlein & Fraefel, 2016). Im Studiengang Sekundarstufe I müssen die Studierenden schriftliche Planungsdokumente einer Lektion, ein Unterrichtsvideo dieser Lektion und schriftliche Analysen des Unterrichtsvideos einreichen. Diese Unterlagen werden von je zwei Gutachter*innen¹ – von einer Person aus der Erziehungswissenschaft und einer aus der Fachdidaktik – anhand eines vorgegebenen Rasters beurteilt.

Eine zuverlässige Beurteilung solcher Abschlussprüfungen ist besonders wichtig, denn die Note kann eine reale Konsequenz haben: Sie kann für die Stellensuche bedeutend sein (Strietholt & Terhart, 2009). Das gilt auch für das Video-Portfolio, denn es ist der einzige benotete Leistungsnachweis des Studienbereichs *Berufspraktische Studien* und nimmt im Diplomzeugnis somit einen hohen Stellenwert ein.

In diesem Beitrag liegt der Fokus auf der Validität der Überprüfung der schriftlichen Analysekompetenzen im Rahmen des Video-Portfolios.

Analysekompetenz als Bestandteil der Reflexion

Reflexion kann als ein kognitiver Prozess beschrieben werden, der sich auf einen Gegenstand bezieht und ein gewisses Ziel hat. Durch diesen Gedankenprozess werden mentale Strukturen geschaffen oder verändert, die das eigene Verhalten (mit-)steuern (Aeppli & Lötscher, 2016). Die Reflexion des eigenen Unterrichts gilt heute als zentrale Kompetenz von Lehrpersonen (Häcker, 2017). Sie trägt zur Professionalisierung bei, weil sie persönliche und professionelle Entwicklung ermöglicht (Wyss, 2013). Die Analyse ist eine Komponente der Reflexion, bei

1 Die Gutachter*innen sind Ausbilder*innen, die in verschiedenen Ausbildungsberufen der PH FHNW beschäftigt sind.

der es um eine vertiefte Auseinandersetzung mit einem Gegenstand geht (Aeppli & Lötscher, 2016) – im Rahmen der Video-Portfolios um Sequenzen des eigenen Unterrichts. In der Forschungsliteratur spricht man auch von der *professionellen Unterrichtswahrnehmung*, bei der (angehende) Lehrpersonen bestimmte Unterrichtssituationen vor dem Hintergrund ihres pädagogischen und fachdidaktischen Wissens beobachten und interpretieren (Jahn, Stürmer, Seidel et al., 2014).

Durch Unterrichtsvideos ist die Förderung der Analysekompetenz besonders gut möglich, weil die Studierenden zeigen können, dass sie nicht nur Professionswissen besitzen, sondern dieses für die Analyse von Unterrichtssituationen nutzen oder Diskrepanzen zwischen eigenen, videografierten Handlungen und ihrem Professionswissen erkennen können (Admiraal, Hoeksma, van de Kamp et al., 2011). Daher werden Unterrichtsvideos zunehmend in der Ausbildung angehender Lehrpersonen verwendet (z. B. Gaudin & Chaliès, 2015). So auch im Studiengang Sekundarstufe I der PH FHNW: Die Studierenden lernen während der Ausbildung, lernrelevante Situationen im Unterricht zu identifizieren, vor dem Hintergrund ihres theoretischen Wissens zu interpretieren und begründete Handlungsalternativen abzuleiten.

Die im Studium aufgebauten Analysekompetenzen werden am Ende des Studiums anhand des Video-Portfolios überprüft. Um die Prüfung gezielt weiterentwickeln zu können, wird diese regelmässig anhand von Studierendenbefragungen evaluiert. Im Fokus dieses Artikels steht die Validität der Prüfung der Analysekompetenz, die die Studierenden im Video-Portfolio mit schriftlichen Analysen des eigenen Unterrichts zeigen müssen.

Die inhaltliche Validität

Die inhaltliche Validität gibt an, inwiefern ein Verfahren misst, „was es zu messen vorgibt“ (Wirtz, 2021). Deshalb wird in diesem Abschnitt ausgeführt, was die Studierenden bezüglich der Analysekompetenz in ihrer Ausbildung lernen und was die PH FHNW überprüft. Denn je mehr sich diese beiden Bereiche überschneiden, desto eher wird von einer inhaltlichen Validität des Video-Portfolios ausgegangen.

Für die Analysen der eigenen Unterrichtsvideos im Rahmen des Video-Portfolios müssen die Studierenden zwei Sequenzen auswählen, bei

denen etwas Wichtiges für das Lernen der Schüler*innen geschieht. Sie können auch eine Sequenz auswählen, die nicht wie geplant verlief, mit der sie nicht zufrieden sind und/oder die für das Lernen der Schüler*innen hinderliche Momente enthält. Anschließend werden die folgenden vier Schritte schriftlich bearbeitet (in Anlehnung an Santagata & Guarino, 2011):

1. Die Sequenz wird im Rahmen der Lektion situiert und daraus ein Thema für die Analyse festgelegt.
2. Zum festgelegten Thema der Analyse werden theoretische und empirische Erkenntnisse aus den Bereichen Fachdidaktik/Erziehungswissenschaft zusammengefasst und daraus Indikatoren abgeleitet und beschrieben, anhand derer die Sequenz analysiert wird.
3. Das Handeln der Lehrperson und dessen Wirkung auf die Schüler*innen werden beschrieben und basierend auf den Indikatoren interpretiert.
4. Aus der Analyse werden Handlungsalternativen abgeleitet, die mit den theoretischen bzw. empirischen Erkenntnissen zusammenhängen.

Die Anwendung dieser Heuristik wird im Rahmen der Ausbildung in verschiedenen Seminaren der Berufspraktischen Studien Sek I geübt. Somit geht das Prüfungskonzept konform mit bestimmten Inhalten des Ausbildungskonzepts. Durch die inhaltliche Kongruenz der Prüfung mit Ausbildungsinhalten wird die inhaltliche Validität der Prüfung gewährleistet (Bäuerlein, Senn & Fraefel, 2019).

Die ökologische Validität

Im Rahmen von Beurteilungsverfahren, die praxis- bzw. berufsnah sind, ist die ökologische Validität von Bedeutung. Denn sie bezeichnet, inwiefern eine (Prüfungs-)Situation mit der Wirklichkeit übereinstimmt (Fahrenberg, 2021). Die ökologische Validität ist für die PH FHNW relevant, da die berufspraktischen Kompetenzen der Studierenden möglichst realitätsnah beurteilt werden sollten, bevor die Absolvent*innen in ein (unbefristetes) Arbeitsverhältnis als Lehrperson an einer Schule eintreten können.

Um die ökologische Validität eines Verfahrens zu bestimmen, können Selbsteinschätzungen beigezogen werden (Gröschner & Schmitt,

2012). In den unten beschriebenen Daten handelt es sich um solche Selbsteinschätzungen: Es sind Angaben von Studierenden im Rahmen einer regelmäßigen Evaluation des Video-Portfolios.

Basierend auf den Annahmen, dass die inhaltliche Validität gewährleistet ist und mittels eines Fragebogens die ökologische Validität näher gefasst werden kann, wird in diesem Beitrag folgende Fragestellung adressiert: Inwiefern schätzen die Studierenden das Verfahren zur Beurteilung der Analysekompetenz als ökologisch valide ein?

Methode

Die Fragestellung wird anhand der Evaluationsdaten von $N = 51$ Studierenden aus drei Semestern (Herbst 2019 bis Herbst 2020) beantwortet (Rücklauf: 50.5%). Der anonyme Online-Fragebogen wird den Studierenden jeweils nach Erhalt der Beurteilung des Video-Portfolios zugestellt. Der Fragebogen enthält vier geschlossene Items zur ökologischen Validität des Analyseteils der Prüfung sowie Items zu den Rahmenbedingungen, zur Organisation und zur Ausbildungssituation. Alle Items haben eine vierstufige Antwortskala, der Wortlaut der Aussagen zur ökologischen Validität kann der Tabelle 1 entnommen werden. Die quantitativen Daten wurden deskriptiv ausgewertet.

Ergebnisse

Tabelle 1 zeigt die Häufigkeitsverteilungen der Antworten der Studierenden zu den validitätsbezogenen Items des Evaluationsfragebogens. Mehr als 90% der Studierenden stimmten der Aussage (eher) zu, dass sie im Video-Portfolio ihre Kompetenzen zeigen konnten, eigene Unterrichtsvideos zu analysieren und daraus Schlussfolgerungen für ihr künftiges Handeln zu ziehen. Zudem fanden es 78% der Studierenden (eher) zutreffend, dass sich Unterrichtsvideos für die Unterrichtsanalyse eignen.

Ferner stimmten 88% der Studierenden der Aussage (eher) zu, dass sie anhand der Unterrichtsvideos Stärken und Schwächen ihres eigenen Unterrichts identifizieren konnten. Das Video-Portfolio ermöglicht somit eine Selbsteinschätzung des aktuellen Kompetenzstands und scheint für die Studierenden selbst wichtige Erkenntnisse bezüglich

ihres Leistungsstands zu bieten. Bezüglich der Frage, ob das Wissen, das sich die Studierenden bei der Arbeit am Video-Portfolio angeeignet haben, auch in der Praxis brauchbar ist, fallen die Antworten weniger eindeutig aus. Knapp 46% der Studierenden sind der Meinung, dass dies (eher) nicht der Fall ist.

Tab. 1 Häufigkeitsverteilungen der Antworten der Studierenden zu den validitätsbezogenen Items im Evaluationsfragebogen

Item	Trifft nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft zu
Im Video-Portfolio konnte ich zeigen, dass ich meinen Unterricht analysieren und Schlussfolgerungen für mein künftiges Handeln ableiten kann.	2.0%	6.0%	42.0%	50.0%
Die Unterrichtsvideos waren für mich ein geeignetes Mittel, um meinen Unterricht zu analysieren.	4.0%	18.0%	34.0%	44.0%
Dank der Arbeit mit meinen eigenen Unterrichtsvideos konnte ich Stärken bzw. Schwachstellen in meinem Unterricht erkennen.	0.0%	12.5%	45.8%	41.7%
Wissen, das ich mir bei der Arbeit am Video-Portfolio angeeignet habe, kann ich in der Praxis anwenden.	10.4%	35.4%	35.4%	18.8%

Insgesamt fallen die Einschätzungen der Studierenden bezüglich der Validität des Analyseteils des Video-Portfolios somit grundsätzlich positiv aus. Ein Großteil der Studierenden scheint der Meinung zu sein, dass ihre Analysekompetenzen in diesem Prüfungsformat angemessen abgebildet werden.

Diskussion

Die Ergebnisse der Studierendenbefragung sprechen insgesamt dafür, dass mit dem Analyseteil des Video-Portfolios das beurteilt werden kann, was intendiert ist. Aus bisherigen Studien ist bekannt, dass die Arbeit mit eigenen Videos besser akzeptiert wird als die Arbeit mit

fremden (Krammer, Hugener, Biaggi et al., 2016). Dies könnte auch auf die Video-Portfolios zutreffen und ein Grund für die insgesamt positiven Einschätzungen der Studierenden sein. Jedoch ist die Rückmeldung, dass das Wissen in der Praxis nur in ca. 50% der Fälle (eher) angewendet werden kann, in Bezug auf die ökologische Validität nicht zufriedenstellend.

Ein möglicher Grund für diese eher kritische Einschätzung könnte sein, dass die Analyse im Rahmen einer Prüfung erstellt und benotet wird. Dies ist ein grundsätzliches Strukturproblem (Klewin, Köker & Störtländer, 2020) und kann zu einer zurückhaltenden oder gar defensiven Reflexion führen, was auch den Transfer erschweren oder behindern kann. Daneben können auch Effekte sozialer Erwünschtheit nicht ausgeschlossen werden (Häcker, 2017). Dieses Grundsatzproblem zu lösen, ist schwierig, und kann vermutlich nicht gänzlich aufgelöst werden. Hilfreich könnte sein, dass die Studierenden in weiteren Seminaren mit Videoanalysen und darauf bezogenen Reflexionsgesprächen arbeiten, damit sich Routinen etablieren, die sich allenfalls positiv auf den Transfer auswirken.

Insgesamt kann dem Prüfungsformat Video-Portfolio aufgrund der Erkenntnisse eine hohe inhaltliche und eine mittlere ökologische Validität attestiert werden. Es ist jedoch zu beachten, dass es bei diesem Prüfungsformat verschiedene Einflussvariablen gibt, die die Reliabilität und Validität der Beurteilung beeinflussen können:

- Die Beurteilung der Video-Portfolios anhand des Rasters verlangt von den Beurteilenden hoch inferente Urteile. Das kann mögliche Urteilsfehler begünstigen (Lotz, Gabriel & Lipowsky, 2013). Deshalb werden die Gutachter*innen an der PH FHNW geschult und das gemeinsame Verständnis an die Anforderungen zu schärfen versucht.
- Qualitativ gute Aufnahmen zu machen, ist eine weitere Herausforderung: Auf den Videoaufnahmen der Studierenden ist stellenweise nicht alles sicht- und hörbar, was es für die Studierenden schwierig machen kann, geeignete Sequenzen für die Analyse zu finden.

Die vorliegenden Daten sind Selbsteinschätzungen, deren Aussagekraft limitiert ist. Zusätzlich kann die Note, die die Studierenden vor Erhalt der Evaluation erhalten haben, die Einschätzungen beeinflussen. Die tatsächliche Wirkung der Videoanalysen auf das unterricht-

liche Handeln der Studierenden ist bislang weitgehend unbekannt (Gaudin & Chaliès, 2015).

Die Evaluation von Abschlussprüfungen wie dem Video-Portfolio entlang der Gütekriterien Reliabilität und Validität wie auch Objektivität und Ökonomie ist zentral. Denn nur so können Hochschulen möglichst transparente und faire Beurteilungen gewährleisten. Gleichzeitig ermöglicht die Prüfung, dass institutsintern über die Ausbildung gesprochen und diese gezielt weiterentwickelt wird.

Literatur

- Admiraal, W., Hoeksma, M., van de Kamp, M.-T. & van Duin, G. (2011). Assessment of teacher competence using video portfolios: Reliability, construct validity, and consequential validity. *Teaching and Teacher Education*, 27, 1019-1028.
- Aeppli, J. & Lötscher, H. (2016). EDAMA – ein Rahmenmodell für Reflexion. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 34, 78-97.
- Bäuerlein, K. & Fraefel, U. (2016). Ein Video-Portfolio zur summativen Überprüfung berufspraktischer Kompetenzen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 34(2), 212-218.
- Bäuerlein, K., Senn, S. & Fraefel, U. (2019). Die Beurteilung berufspraktischer Kompetenzen am Ende der Lehrpersonen ausbildung in der Deutschschweiz. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 37(3), 401-429.
- Esslinger-Hinz, I. (2016). *Gut vorbereitet in die Lehrprobe*. Weinheim: Beltz.
- Fahrenberg, J. (2021). *Ökologische Validität*. Verfügbar unter <https://dorsch.hogrefe.com/stichwort/oekologische-validitaet> [13.12.2021].
- Gaudin, C. & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, 16, 41-67.
- Gröschner, A. & Schmitt, C. (2012). Kompetenzentwicklung im Praktikum? *Lehrerbildung auf dem Prüfstand*, 5(2), 112-128.
- Jahn, G., Stürmer, K., Seidel, T. & Prenzel, M. (2014). Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46(4), 171-180.
- Häcker, T. (2017). Grundlagen und Implikationen der Forderung nach Förderung von Reflexivität in der Lehrerinnen- und Lehrerbildung. In C. Berndt, T. Häcker & T. Leonhard (Hrsg.), *Studien zur Professionsforschung und Lehrerbildung. Reflexive Lehrerbildung revisited: Traditionen – Zugänge – Perspektiven* (S. 21-45). Bad Heilbrunn: Klinkhardt.
- Klewin, G., Köker, A. & Störfländer J. G. (2020). Veranlasste und benotete Reflexion: ein unmögliches Prüfungsformat? *HLZ*, 3(2), 108-121.
- Krammer, K. Hugener, I., Biaggi, S., Frommelt, M., Furrer, G. & auf der Maur, K. (2016). Videos in der Ausbildung von Lehrkräften. *Unterrichtswissenschaft*, 44(4), 357-372.
- Lotz, M., Gabriel, K. & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. *Zeitschrift für Pädagogik*, 3, 357-380.
- Santagata, R. & Guarino, J. (2011). Using video to teach future teachers to learn from teaching. *ZDM*, 43, 133-145.

- Strietholt, R. & Terhart, E. (2009). Referendare beurteilen. Eine explorative Analyse von Beurteilungsinstrumenten in der Zweiten Phase der Lehrerbildung. *Zeitschrift für Pädagogik*, 55(4), 622-645.
- Wirtz, M. A. (2021). *Inhaltliche Validität*. Verfügbar unter <https://dorsch.hogrefe.com/stichwort/validitaet-inhaltliche> [13.12.2021].
- Wyss, C. (2013). Unterricht und Reflexion: eine mehrperspektivische Untersuchung der Unterrichts- und Reflexionskompetenz von Lehrkräften. *Empirische Erziehungswissenschaft*, Band 44. Münster: Waxmann.

David Bisang, wiss. Mitarbeiter
an der Professur für Berufspraktische Studien
und Professionalisierung Sek I
der PH FHNW.
Arbeitsschwerpunkte:
Video-Portfolio

david.bisang@fhnw.ch



Kerstin Bäuerlein, Dr., Dozentin
an den Professuren für Berufspraktische Studien
und Professionalisierung Sek I und II
der PH FHNW.
Arbeitsschwerpunkte:
digitale Medien, selbstreguliertes Lernen

kerstin.baeuerlein@fhnw.ch



Corinne Wyss, Prof. Dr., Leiterin
der Professur für Berufspraktische Studien
und Professionalisierung Sek I
der PH FHNW.
Arbeitsschwerpunkte:
Professionalisierung,
Reflexion, digitale Medien

corinne.wyss@fhnw.ch



03

Christoph Kruse

Die Beurteilung angehender
Lehrer*innen im Referendariat.
Relevanz, Kontext
und erste Explorationen

Die Personalbeurteilung in der Lehrer*innenbildung kann generell als hoch relevantes und zugleich wenig erforschtes Thema angesehen werden (König, 2020). Dies gilt insbesondere für die zweite Phase, für das Referendariat, ist hier doch ein – verglichen mit der ersten Phase – breites Spektrum des Lehrer*innenhandelns Gegenstand der Beurteilung. In diesem Beitrag werden zunächst die Relevanz der Beurteilungen von Lehrer*innen im Vorbereitungsdienst (LiV) betrachtet und Rahmenbedingungen dieser Beurteilungen am Beispiel des deutschen Bundeslandes Nordrhein-Westfalen (NRW) dargelegt. Darauf aufbauend wird ein Forschungsprojekt vorgestellt, das sich der Beurteilung im Referendariat annimmt und schriftliche Gutachten von Fach- und Schulleitungen zum Forschungsgegenstand macht. Im Anschluss an einen ersten Einblick werden weitere Forschungsperspektiven und Folgefragen abgeleitet. In diesem Beitrag sollen also weder einzelne Forschungsbefunde, der Beurteilungskontext noch eine exemplarische Gutachtenpassage in ihrem Interpretationspotenzial vollumfänglich diskutiert werden. Stattdessen ist es gemäß der vorgestellten theoriegenerierenden qualitativen Studie das Ziel, aufbauend auf der Bedeutung und dem Beurteilungskontext erste Eindrücke und Spuren der Analyse vorzustellen.

Relevanz

Die Bedeutung von Beurteilungen angehender Lehrer*innen im Vorbereitungsdienst (VD) wird im Folgenden aus drei Perspektiven hergeleitet: Zum einen handelt es sich bei dem in der zweiten Phase erreichten Abschluss des zweiten Staatsexamens in Deutschland um die zentrale Zugangsvoraussetzung zum Lehrer*innenberuf. Die Relevanz der hiermit verbundenen Selektionsfunktion der Beurteilung wird zudem noch dadurch verstärkt, dass es in Deutschland (mit Ausnahme des Bundeslandes Bayern) nach erfolgter Festanstellung/Verbeamtung keine fest verankerte Beurteilungspraxis zu einem weiteren, späteren Zeitpunkt gibt (OECD, 2018; Terhart, 2016). Es handelt sich bei der Beurteilung im VD folglich um die wichtigste institutionalisierte Beurteilung des umfassenden Handlungsspektrums von Lehrer*innen im Berufsfeld. Nachgelagerte reguläre und formalisierte Beurteilungen existieren lediglich im Rahmen der Ernennung zum*zur Beamt*in auf Lebenszeit (sowie bei vereinzelt Bewerbungen auf Beförderungs-

stellen; vgl. Tulowitzki & Kruse, 2020). Welchen Stellenwert diese weiteren, formal geforderten Beurteilungen im Dienstverhältnis realiter erhalten, wurde bislang nicht erforscht. Kurzgefasst: Die Beurteilung im zweiten Staatsexamen und somit die Selektion in den Beruf hinein erhält eine erhöhte Bedeutung, da innerhalb des Berufes kaum noch beurteilt, geschweige denn selektiert wird.

Zweitens lässt sich annehmen, dass die LiV aufgrund der somit hohen Selektionsrelevanz des zweiten Staatsexamens ihre Lern- und Entwicklungs- bzw. Professionalisierungsbemühungen auf die zu beurteilenden Aspekte konzentrieren. Insofern haben wir es bei den Beurteilungen in der zweiten Phase mit bedeutsamen Ausbildungsimpulsen in der Lehrer*innenbildung zu tun.

Und drittens lässt sich die Bedeutung der Beurteilung in der zweiten Phase empirisch belegen: Das Beurteilungsergebnis des zweiten Staatsexamens hat gemäß Roloff, Klusmann, Lüdtke und Trautwein (2020) einen (signifikant) positiven Einfluss auf die Unterrichtsqualität. Dem Beurteilungsergebnis der ersten, universitären Phase der Lehrer*innenbildung kommt hier bemerkenswerterweise kein bzw. ein tendenziell negativer (nicht signifikanter) Einfluss auf die vier erhobenen Unterrichtsqualitätsmerkmale zu (Roloff et al., 2020). An dieser Stelle bleibt zunächst das empirisch belegbare Argument der prädiktiven Validität der Beurteilungspraxis im Referendariat hinsichtlich der Unterrichtsqualität festzuhalten: Die Beurteilung im VD ist folglich (auch) deshalb bedeutsam und wissenschaftlich interessant, weil sie ein (zumindest auf o. g. empirischer Basis in Teilen) valides Instrument zur Vorhersage des Berufserfolgs darstellt.

Aus dieser Bedeutung für die Lehrer*innenbildung und den Lehrer*innenberuf heraus ist zunächst zu fragen: Wie ist die Beurteilung im VD überhaupt organisiert und welche Akteure sind involviert?

Rahmenbedingungen der Beurteilungspraxis

Grundlage der folgenden Ausführungen ist die zentrale Verordnung zum VD in NRW (OVP, 2022)¹.

¹ Anpassungen dieser Verordnung an die besondere Situation von Schulen in der Corona-Pandemie wurden für diesen Beitrag nicht berücksichtigt.

Grundsätzlich setzt sich die Abschlussnote der LiV aus zwei gleich gewichteten Bereichen zusammen: Die Beurteilung der Leistungen am Prüfungstag und die den Verlauf des VD insgesamt betreffende Beurteilung mittels schriftlicher Gutachten (Abb. 1).

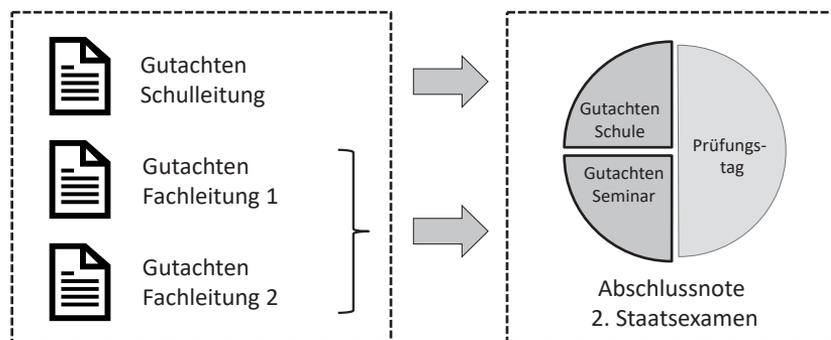


Abb. 1 Schriftliche Gutachten und Abschlussnote im zweiten Staatsexamen in NRW

Beurteilt werden am Prüfungstag zwei Unterrichtsdurchführungen und sich jeweils anschließende Reflexionsgespräche, die zugehörigen schriftlichen Planungsentwürfe und ein übergreifendes Prüfungsgespräch zu „zentralen Bereichen des beruflichen Handelns“ (OVP, 2022, § 33 Abs. 2). Die Prüfungskommission besteht aus einer Schulleitung und zwei Fachleitungen: Ausbilder*innen für die jeweiligen Unterrichtsfächer des*der LiV an Seminarstandorten (Zentren für schulpraktische Lehrerausbildung, ZfsL, ehem. Studienseminare). Eine dieser Fachleitungen wird von dem*der LiV benannt und hat die anderthalbjährige Ausbildung begleitet. Die anderen beiden Mitglieder der Prüfungskommission kennen den*die LiV vorab nicht.

Die andere Hälfte der Gesamtnote des zweiten Staatsexamens setzt sich aus schriftlichen Beurteilungen von den an der Ausbildung beteiligten Akteuren zusammen (Abb. 1). Verfasst werden diese Gutachten von der Schulleitung der jeweiligen Ausbildungsschule und von den beiden Fachleitungen, die die Ausbildung des*der LiV am ZfsL verantworten und gestalten. Die Note der beiden Gutachten der Fachleitungen wird i. d. R. arithmetisch gemittelt und geht zusammengekommen in die Gesamtbeurteilung ein. Beim Verfassen gilt es, „die fachlichen als auch die überfachlichen Kompetenzen auf der Grundlage von in Ausbildungssituationen aller Art gemachten Beobachtungen

in ihrer Gesamtheit zu bewerten“ (OVP, 2022, § 16 Abs. 2). Hierzu gehören insbesondere (i. d. R.) je fünf Unterrichtsbesuche pro Fach, die – analog zum Prüfungstag – von einem schriftlichen Planungsentwurf und einer anschließenden Reflexion flankiert werden. Auch die Schulleitung der Ausbildungsschule kann an diesen Besuchen teilnehmen. Ferner fließen in das Schulleitungsgutachten weitere schriftliche Beurteilungen von den an der Ausbildung beteiligten Lehrer*innen des Kollegiums ein. Formale Grundlage aller Gutachten und Beurteilungen sind stets die (bildungswissenschaftlichen) Standards für die praktischen Ausbildungsabschnitte sowie die übergeordneten Kompetenzen gemäß der KMK (2019a; vgl. OVP, 2022, insb. Anlage 1). Das zeigt sich auch darin, dass die seitens der KMK definierten Handlungsfelder und Kompetenzen zentrale Aspekte der formalen Gutachtenvorlage darstellen und die Beurteilenden somit aufgefordert sind, ihre schriftlichen Begründungen an der KMK-Vorlage auszurichten. In den ersten Explorationen in diesem Beitrag kann dies exemplarisch anhand einer Gutachtenpassage verdeutlicht werden (s. u. Abb. 2).

Beispielhaft sollen folgende Aspekte der Rahmenbedingungen in den Blick genommen werden: Zum einen wirkt die Definition „Ausbildungssituationen aller Art“ (OVP, 2022, § 16 Abs. 2) als Beurteilungsgrundlage diffus und umfassend: Lässt sich diese Formulierung bereits als Indiz dafür lesen, dass *jedliches* Verhalten der LiV zum Beurteilungsgegenstand werden kann, da es sich prinzipiell den betreuenden (und ja auch beurteilenden) Lehrer*innen sowie den Fach- und Schulleitungen gegenüber per se um ein Ausbildungsverhältnis handelt? Und zweitens erscheint bedeutsam, dass gerade die *bildungswissenschaftlichen* Kompetenzen und zugehörigen Standards in der Beurteilung eine formal so prominente Relevanz erhalten. Wie steht es folglich in der Beurteilungspraxis um die *fachlichen und fachdidaktischen* Kompetenzen, zu denen es ebenfalls (wenngleich auch nicht durchgängig für den VD konkretisierte) Anforderungen seitens der KMK (2019b) gibt? Übergeordnet und grundsätzlich stellt sich die Frage danach, auf welche Weise die Beurteilenden auf die formal verankerte Beurteilungsaufforderung reagieren: Welche Kriterien und Begründungen werden in der Beurteilungspraxis tatsächlich (in Übereinstimmung mit der bildungswissenschaftlichen KMK-Vorlage (2019a) oder über diese hinausgehend) angeführt?

Diese Aspekte verdeutlichen beispielhaft, dass die Kenntnis formaler Rahmenbedingungen der Beurteilungspraxis aus wissenschaftlicher

Perspektive mitunter mehr Fragen als Antworten provoziert. Das Problem eines „unzureichenden Konsens über die Erfolgskriterien zur Beurteilung einer gelungenen Berufseinführungsphase im wissenschaftlichen und politischen Diskurs“ (Rauin, 2014, S. 575) scheint folglich trotz (bildungswissenschaftlicher) KMK-Vorlage weiterhin ungeklärt. Denn schließlich ist aus wissenschaftlicher Perspektive, von wenigen Ausnahmen (z. B. Kärner, Bonnes & Schölzel, 2019; Strietholt & Terhart, 2009; Wernet, 2009) abgesehen, die Beurteilungspraxis in der zweiten Phase nahezu unerforscht und daher unklar, welche Kriterien in der Praxis tatsächlich greifen. Entsprechend gilt auch im aktuellen wissenschaftlichen Diskurs das Desiderat bezüglich „fachlicher Erwartungshaltungen im VD“ sowie zu „Erwartungshaltungen von [...] beurteilenden Schulleitungen, die als wichtige Akteure im VD fungieren“ (Reintjes, Bellenberg & Labott, 2021, S. 128).

Forschungsprojekt zu schriftlichen Gutachten

Dieser Desiderata nimmt sich das vorzustellende Forschungsprojekt mit dem Ziel an, die Beurteilungspraxis im VD anhand schriftlicher Gutachten rekonstruktiv zu erforschen. Vermittelt durch eine solche Rekonstruktion werden gleichsam die verschiedenen Kriterien, Normen und Idealvorstellungen analysiert, die der Bewertung zugrunde liegen. Übergeordnete Fragestellungen lauten: Welche Erwartungen im Sinne eines Ausbildungsideals am Ende des Referendariats werden in den Beurteilungen deutlich? Welche Gründe für vorhandene und nicht vorhandene Fähigkeiten werden rekonstruierbar und wie wird das Vorhandensein/Fehlen dieser Fähigkeiten in den Gutachten erklärt?

Um diesen Fragestellungen fundiert und umfassend nachzugehen, konnte mit Unterstützung der zuständigen Landesbehörde für das Bundesland NRW kriteriengeleitet ein anonymisierter Datensatz zusammengestellt werden, anhand dessen eine möglichst große Varianz der schriftlichen Beurteilungspraxis in den Blick genommen werden kann. Berücksichtigte Kriterien waren hierbei das Ausbildungslehramt (Primar-, sowie Sekundarstufe I und II), die Autorenschaft der Beurteilung (Fach- bzw. Schulleitung), das Schulfach sowie das Beurteilungsergebnis. Zur Auswertung wird sich im Rahmen einer explorativen Dokumentenanalyse (Hoffmann, 2018) auf die Grounded Theory

Methodologie (GTM; Strauss, 1998) bezogen. Diese Kombination ermöglicht sowohl die umfassende rekonstruktive Analyse zahlreicher Gutachten im Rahmen des theoretical samplings der GTM als auch die systematische Berücksichtigung des Entstehungskontexts der Gutachten, wie sie in einer Dokumentenanalyse gefordert ist und angesichts o. g. Ausführungen und Fragen auch als gegenstandsangemessen erscheint. Erste Einblicke in das Projekt anhand einer ausgewählten Gutachtenpassage können im Folgenden dargestellt werden.

Erste Einblicke anhand einer Gutachtenpassage

Kompetenz 6:

Lehrerinnen und Lehrer finden Lösungsansätze für Schwierigkeiten und Konflikte in Schule und Unterricht.

Frau (LiV) tritt den Schülerinnen und Schülern offen, zugewandt und hilfsbereit gegenüber; Im Umgang mit Ihnen ist sie immer freundlich. Ihr Auftreten Schülerinnen und Schülern gegenüber war jedoch manchmal noch von Unsicherheit und Nervosität geprägt, sodass sich dies auf die Lernenden übertrug. Im Laufe der Ausbildungszeit gelang es ihr immer besser, auf Störung und unangebrachtes Verhalten adäquat zu reagieren und entsprechend zu handeln. In einige Situation sollte Frau (LiV) bestimmter reagieren, um Konflikte oder aufkommende Unruhe im Unterricht bereits frühzeitig zu unterbinden. Ihre Authentizität litt manchmal noch unter ihrer Inkonsequenz und Unsicherheit.

Abb. 2 Passage eines schriftlichen Gutachtens

Es geht am Beispiel dieser Passage nun vornehmlich darum, „generative Fragen“ (Strauss, 1998, S. 44) als erste Annäherungen an das Datenmaterial zu formulieren. Diese Fragen sind im Rahmen der (theoriegenerierenden) GTM ein wichtiges Instrument zur weiterführenden Analyse, denn sie bilden die am Material entwickelte zunehmende Schärfung der Forschungsperspektive ab. Die nun illustrierten ersten Spuren und Folgefragen zur Praxis der schriftlichen Beurteilungslegitimation stellen somit noch keine Entwicklung von Kodes oder Kategorien (wie in der GTM anvisiert) dar, sondern sind als Vorstufe zu einem solchen Schritt zu verstehen.

Ganz grundsätzlich lässt sich die Passage hinsichtlich verschiedener Perspektiven interpretieren. Eine dieser Perspektiven betrifft die *Strategien der Bewertung*: Wie bearbeitet der*die Verfasser*in die An-

forderung, entsprechende Wertungen vorzunehmen? Es deutet sich, zunächst unabhängig von einem konkret zu beurteilenden Aspekt des Lehrer*innenhandelns, eine dichte Verschränkung verschiedener Bewertungsstrategien an: Defizitkonstruktionen und Kritik werden mit positiven Konnotationen (erster Satz), Entwicklungsperspektiven („manchmal noch“), Entwicklungsbeschreibungen („gelang es ihr immer besser“) sowie (Entwicklungs-)Erwartungen („sollte [...] bestimmter reagieren“) verbunden. Daraus ergeben sich weitere Fragen: Zeigt sich eine solche Verschränkung durchgängig und welche zusätzlichen Wertungsstrategien zeigen sich insgesamt im Material? In welchem Verhältnis stehen diese zu den jeweils spezifisch beurteilten Berufsaspekten und auch zu der abschließenden Beurteilung in Form einer Note? Lassen sich diese Strategien als Indizien zu Entwicklungsverständnissen, etwa im Sinne von Professionalisierungsannahmen lesen? Spezifischer ließe sich auch fragen: In welchem Verhältnis stehen die so rekonstruierten Bewertungen zu der Annahme, dass am Ende der zweiten Phase die beruflichen Fähigkeiten nicht vollends ausgeprägt, sondern auch nur grundgelegt werden können?

Als Interpretation in einer weiteren Perspektive wird eine Strategie ersichtlich, wie die KMK-Kompetenzen als Beurteilungsmaßstab in den Gutachten genutzt werden: Der*die Autor*in isoliert eine einzelne Kompetenz und formuliert dazu einen Absatz. Die Lesenden würden nun eine Passage erwarten, in der der Grad der Kompetenzerreichung der LiV zur „Kompetenz 6“ beurteilt wird.

Welcher Aspekt des Lehrer*innenhandelns wird unter dem abstrakten Finden von „Lösungsansätzen für Schwierigkeiten und Konflikte[n]“ nun als Beurteilungskriterium fokussiert? Es scheint um den Umgang mit Unterrichtsstörungen, Konflikten und Unruhe als Teilaspekte von Klassenführung zu gehen. Dass die beurteilte LiV hier tendenziell hinter den Erwartungen zurückbleibt, ist schnell ersichtlich. Interessant ist indes, wie diese Beurteilung begründet wird: Nicht nur fehlende Fähigkeiten oder Techniken der Klassenführung („adäquat zu reagieren“) werden als Argument angeführt, sondern insbesondere „Unsicherheit und Nervosität“ im „Auftreten Schülerinnen und Schülern gegenüber“ scheint die Erklärung für Defizite im Lehrer*innenhandeln zu liefern. Schlussendlich wird der LiV eine („manchmal“) unzureichende „Authentizität“ attestiert. Wobei die Argumentation im letzten Satz weitere Fragen aufwirft: In welchem Zusammenhang steht Authentizität hier mit „Inkonsequenz und Unsicherheit“? Wäre es nicht gera-

de authentisch, sich unsicher zu verhalten, wenn man sich unsicher fühlt? Mit welcher Funktion und in welcher Absicht wird hier also der Authentizitätsbegriff mobilisiert? Könnte sich (insbesondere) der letzte Satz dieser Passage als Indiz dafür lesen, dass zur Erklärung eines Kompetenzdefizits verallgemeinerte, bisweilen diffuse *Eigenschaften der Person*, der „Persönlichkeit“, herangezogen werden?

Zumindest erhalten wir in dieser Passage Hinweise darauf, dass mehr geschieht, als dass die Kompetenzformulierung der KMK lediglich konkretisiert oder hinsichtlich einer individualisierten Beurteilung rekontextualisiert wird. Nicht (nur) der Grad einer Kompetenzerreichung wird beurteilt, sondern das „Auftreten Schülerinnen und Schülern gegenüber“ als grundsätzliches Verhalten der Person wird zum Beurteilungsgegenstand, wenn über die Klassenführung der LiV geurteilt wird.

Diskussion und Ausblick

Eine sich hier andeutende Beurteilungslegitimation wäre für die Lehrer*innenbildung bemerkenswert und differenziert zu diskutieren: Wird der Kompetenzbegriff verwendet, um Berufsfacetten zu beschreiben und beurteilen, so kann von dessen grundsätzlicher Erlernbarkeit ausgegangen werden – so auch bezogen auf die hier in Teilen verhandelte Klassenführungskompetenz. Mit Begriffen wie „Nervosität“, „Unsicherheit“ oder „Authentizität“ werden hingegen (partiell) Persönlichkeitsaspekte zum Gegenstand der Beurteilung und damit für die Erklärung des Berufserfolgs herangezogen, die entweder grundsätzlich nicht erlernt oder zumindest nicht im anderthalbjährigen Referendariat verändert resp. erworben werden können (zur Problematisierung der *Lehrerpersönlichkeit* in der Lehrer*innenbildung siehe Rothland, 2021). Wird also im Referendariat etwas beurteilt, was dort gar nicht vermittelt bzw. erworben werden kann oder grundsätzlich nicht lehr-, nicht bestimm- und nicht erlernbar erscheint? Aus welcher Perspektive wäre dies ein problematischer Befund und welche Funktion erfüllt eine solche Beurteilungsstrategie für die Beurteilenden? Wären es, in Analogie zu Rothlands (2021) Ausführungen zum wissenschaftlichen Diskurs, die Diffusität und Unbestimmbarkeit einer *Lehrerpersönlichkeit*, die sie in der Beurteilungspraxis als Argument funktional erscheinen lassen? Oder ließe sich das Konstrukt durch systematische Rekonstruktionen anhand der Gutachten empirisch spe-

zifischer fassen? Übergeordnet wird dementsprechend auch zu diskutieren sein, in welchem Verhältnis die durch die Beurteilungspraxis vermittelten Ausbildungsideale und Entwicklungsannahmen zu den im wissenschaftlichen Diskurs verhandelten Professionalitäts- und Professionalisierungsansätzen stehen.

Ob die anhand der ausgewählten Passage beschriebenen ersten Hinweise für durchgängig rekonstruierbare Bewertungs- und Argumentationsstrategien stehen, müssen systematischere Auswertungen im Rahmen des Projekts erst zeigen. Denkbar wäre ebenfalls, dass zu verschiedenen Aspekten des Lehrer*innenhandelns unterschiedliche Beurteilungsstrategien herangezogen werden. Welche Berufsfacetten überhaupt Gegenstand der Beurteilung sind, auf welche Weise sie bewertet werden und in welches Verhältnis diese Facetten zur (bildungswissenschaftlichen) KMK-Vorlage gebracht werden, sind angesichts der herausgestellten Relevanz der Beurteilung im VD weitere im Forschungsprojekt zu bearbeitende Fragen.

Literatur

- Hoffmann, N. (2018). *Dokumentenanalyse in der Bildungs- und Sozialforschung. Überblick und Einführung*. Weinheim: Beltz Juventa.
- Kärner, T., Bonnes, C. & Schölzel, C. (2019). Bewertungstransparenz im Referendariat. *Zeitschrift für Pädagogik*, 65(3), 378-400.
- KMK (2019a). *Standards für die Lehrerbildung: Bildungswissenschaften* (Fassung vom 16.05.2019). Sekretariat der Kultusministerkonferenz.
- KMK (2019b). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung* (Fassung vom 16.05.2019). Sekretariat der Kultusministerkonferenz.
- König, J. (2020). Beurteilung und Zertifizierung von (angehenden) Lehrerinnen und Lehrern. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 376-384). Bad Heilbrunn: Klinkhardt.
- OECD (2018). *Effective Teacher Policies: Insights from PISA*. Paris: OECD.
- OVP (2022). *Ordnung des Vorbereitungsdienstes und der Staatsprüfung für Lehramter an Schulen (OVP)*. RECHT.NRW.DE. Verfügbar unter https://recht.nrw.de/lmi/owa/br_text_anzeigen?v_id=10000000000000000681 [04.01.2022].
- Rauin, U. (2014). Forschung zur Lehrerbildung aus nationaler und internationaler Perspektive. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2., überarb. und erw. Aufl.) (S. 572-579). Münster u. a.: Waxmann.
- Reintjes, C., Bellenberg, G. & Labott, D. (2021). Erwartungen an Lehrkräfte im Vorbereitungsdienst. Eine Bestandsaufnahme. In J. Peitz & M. Harring (Hrsg.), *Das Referendariat. Ein systematischer Blick auf den schulpraktischen Vorbereitungsdienst* (S. 119-129). Münster u. a.: Waxmann.

- Roloff, J., Klusmann, U., Lüdtke, O. & Trautwein, U. (2020). The Predictive Validity of Teachers' Personality, Cognitive and Academic Abilities at the End of High School on Instructional Quality in Germany: A Longitudinal Study. *AERA Open*, 6(1), 1-17. <https://doi.org/10.1177/2332858419897884>
- Rothland, M. (2021). Die „Lehrerpersönlichkeit“: das Geheimnis des Lehrberufs? *Die Deutsche Schule*, 113(2), 188-198. <https://doi.org/10.31244/dd.2021.02.06>
- Strauss, A. L. (1998). *Grundlagen qualitativer Sozialforschung* (2. Aufl.). Paderborn: Fink.
- Strietholt, R. & Terhart, E. (2009). Referendare beurteilen. Eine explorative Analyse von Beurteilungsinstrumenten in der Zweiten Phase der Lehrerbildung. *Zeitschrift für Pädagogik*, 55(4), 622-645.
- Terhart, E. (2016). Personalauswahl, Personaleinsatz und Personalentwicklung an Schulen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (2. Aufl.) (S. 279-299). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-18942-0_10
- Tulowitzki, P. & Kruse, C. (2020). Qualifikation für Schulleitung und besondere Aufgaben. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 360-368). Bad Heilbrunn: Klinkhardt.
- Wernet, A. (2009). Konformismus statt kollegiale Anerkennung: Fallstudien zur Ausbildungskultur im Referendariat am Beispiel von Beurteilungen. *Pädagogische Korrespondenz*, (39), 46-63.

Christoph Kruse, wiss. Mitarbeiter
an der WWU Münster.
Arbeitsschwerpunkte:
Forschung zum Lehrer*innenberuf
und zur Lehrer*innenbildung,
Schulleitungs- und Schulentwicklungsforschung.

christoph.kruse@uni-muenster.de



04

Kerstin Rabenstein

Einblicke in die soziale Praxis
des Prüfens an Hochschulen.
Innen(ein)sichten

Mit dem Umbau der Studienstrukturen in modularisierte Studiengänge hat sich in vielen universitären Fächern die Bedeutung von Prüfungen als die Abschlussphase prägendes Element hin zu einem Element, das sich durch das ganze Studium zieht, entwickelt. Kalthoff und Engert (2021) setzen die Prüfungsdichte in den *alten* Magister- und Diplomstudiengängen bei einer Studiendauer von durchschnittlich zehn Semestern bei fünf bis zehn Prüfungen an, während Bachelorstudierende gegenwärtig schon auf bis zu 30 Prüfungen kommen können. Damit hat die Aufmerksamkeit für Prüfungen im Zuge der Modularisierung des Studiums tendenziell zugenommen. Darüber, wie sich das Prüfen selbst im Hochschulalltag vollzieht und das Lehren und Studieren präfiguriert, wird zwar von Lehrenden und Studierenden mehr oder weniger an offizieller Stelle immer mal wieder diskutiert. Aus der (qualitativen) Forschung wissen wir dazu aber noch wenig.

Die qualitative Forschung interessiert sich nicht für das Ge- oder Misslingen von Prüfungen bzw. die Objektivität der Notengebung. Vielmehr bietet sie durch offene Interviews mit Prüfenden und Geprüften (Kalthoff & Engert, 2021) und audio- bzw. videogestützten in situ Beobachtungen von Prüfungen (Herzmann & Liegmann, 2020; Tyagunova, 2021) bzw. Facebook-Posts (Tyagunova & Greiffenhagen, 2021) Einblicke in die alltäglichen Praktiken des Prüfens als zentrales Element von Studium und Lehre. Die Prüfungspraxis an Hochschulen stellt für die qualitative Forschung ein noch junges und besonders sensibles Forschungsfeld dar. Wenn Prüfungen beobachtet werden, muss sichergestellt sein, dass Prüfungsverläufe und -ergebnisse in keiner Weise gestört und dadurch anfechtbar gemacht werden. Wenn Interviews mit Prüfenden und Geprüften geführt werden, muss für die Anonymisierung der Daten gesorgt werden.

Im Folgenden wird – schlaglichtartig – ein Einblick in aktuelle qualitative Studien zur Prüfungspraxis an Hochschulen gegeben. Da bisher sehr selten explizit Prüfungspraktiken in den lehramtsbezogenen Studiengängen untersucht werden, reflektiert der Beitrag keine Spezifika von Prüfungen in lehramtsbezogenen Studiengängen, sondern allgemeine Entwicklungen. Über die Beschreibung der Innensicht von Prüfenden und Geprüften zielt er auf Einsichten in Bezug auf den Einfluss der gegenwärtigen Prüfungspraxis auf den Alltag des Studierens und Lehrens. Zu Tage treten Spannungen und Ambivalenzen, die den von Prüfungen durchzogenen Lehr- und Studieralltag kennzeichnen.

Wie und von wem wird geprüft?

Die curriculare Rahmung der Lehre ist an den Hochschulen im Rahmen der für die Studiengänge jeweils geltende Gesetzgebung – gegenüber der Praxis an Schulen – etwas loser gehalten: Modulordnungen und Modulhandbücher stecken Kompetenzen und Inhalte ab. Für die ein Modul abschließende Prüfung sehen Modulordnungen die einsetzbaren Prüfungsformen vor, deren Umsetzung dann in der Verantwortlichkeit der Lehrenden liegt. Die Tendenz im Lehramtsstudium geht dabei zum Angebot möglichst vielfältiger Prüfungsformen über das Studium hinweg: Prüfungsformen variieren von Klausuren, die Multiple-Choice- und offene Aufgaben enthalten können, über Präsentationen und mündliche Prüfungen bis hin zu unterschiedlichen Textformaten, wie Hausarbeiten, Portfolios, Essays etc. Wieweit Gruppenprüfungen möglich sind, ist in den Hochschulgesetzen der Länder geregelt. Abschlussprüfungen unterscheiden sich von den laufenden Prüfungen vom Umfang, Anspruch und Gewichtung der Ergebnisse her. Nicht bestandene Prüfungen können unter bestimmten Bedingungen wiederholt werden, die Häufigkeit der Wiederholungsmöglichkeit ist aber begrenzt. Mit Hilfe digitaler Infrastrukturen werden Prüfungsergebnisse den Geprüften angezeigt bzw. von diesen abgerufen. Die Frage, wie Geprüfte über die Ergebnismitteilung hinaus Rückmeldung zu ihren Prüfungsergebnissen erhalten, ist demgegenüber eher wenig geregelt. Zu den wichtigsten Formaten dürften neben persönlichen Gesprächen in Sprechstunden formalisierte Rückmeldungen oder Kurzgutachten gehören.

Zudem macht der Blick auf das Lehrpersonal, das prüft, deutlich, wie heterogen die Voraussetzungen und Bedingungen sind, unter denen geprüft wird (Kalthoff & Engert, 2021). Das Lehrpersonal, das prüft, unterscheidet sich in den lehramtsbezogenen Studiengängen vieler Hochschulen erheblich in Bezug auf den Qualifizierungsgrad und den damit verbundenen Status, die Dauer der Zugehörigkeit zur jeweiligen Institution zwischen befristeten und unbefristeten Verträgen sowie den Umfang der zu verantwortenden Lehre bzw. das Volumen der damit einhergehenden Prüfungen. Einführungsseminare werden bisweilen kaum mehr von Hochschullehrenden gegeben. Für Lehramtsstudiengänge an Universitäten gilt tendenziell ebenfalls: „Die noch in der Qualifikationsphase steckenden, befristeten wissenschaftlichen Mitarbeiter/innen machen den Großteil der wissenschaftlichen Be-

schäftigten an deutschen Hochschulen aus“ (Kalthoff & Engert, 2021, S. 244). Das Lehrpersonal, das sich in einer wissenschaftlichen Qualifikationsphase befindet, ist in der Situation, sich selbst noch Prüfungen für den nächsten Qualifikationsschritt unterziehen zu müssen. In Bezug auf das Lehramtsstudium müssen zudem Lehrbeauftragte und Mentor*innen, denen eine Prüfungsverantwortung zukommt, berücksichtigt werden, also eine Gruppe von Personal, das nicht Mitglied der Universität ist.

Kalthoff und Engert (2021) kommen auf Basis von offenen Interviews mit Prüfenden und Geprüften im Fach ‚Historische Wissenschaften‘ zu dem Schluss:

„Mit dem Abprüfen einzelner Studienmodule geht eine Aufwertung und Steigerung des Prüfens einher, welches die rituelle Logik des Studierens verändert hat. Die Systematik der Bewertung wurde nachhaltig umgebaut von einem konfirmatorischen Prüfungswesen – dem Initiationsritus aus Vordiplom und Diplom, Zwischenprüfung und Magister – zu einer kumulativen Dauerbewertung der Student*innen. Erst gegen Ende, auch bedingt durch die weniger entscheidende Abschlussarbeit, entspannt sich das Prüfungspensum und läuft aus, wengleich die Erwartungshaltungen an die abschließenden Prüfungsleistungen nochmal ansteigen.“ (Kalthoff & Engert, 2021, S. 250)

Diese Umstellung bedeutet aber nicht nur eine Verschiebung der Prüfungslast für den Einzelnen, sondern es zieht auch Re-Konfigurationen der Studienstruktur nach sich. Kalthoff und Engert (2021, S. 250) formulieren für das von ihnen untersuchte Fach: Dies „erfordert für die Geschichtswissenschaft ein Vorverlagern des Stellwerks weg von den *großen Arbeiten* im fortgeschrittenen Studienverlauf hin zu kleinteiligeren Prüfungsmomenten zu Beginn des Studiums.“ Wie sich diese Vorverlagerung des Prüfungsvolumens im universitären Alltag zeigt, wird im Folgenden thematisiert.

Prüfen als Scharnier von Lehren und Lernen

Prüfungen sind als eine Technik zu verstehen, die Prüfende mit einer besonderen Macht ausstatten: Anhand von Prüfungsergebnissen werden Aussagen über das Nicht/Können und Nicht/Wissen des Geprüften glaubhaft gemacht, denen Bedeutung weit über die Prüfungssituation hinaus für Entwürfe und Möglichkeiten der (beruflichen) Zukunft

zugeschrieben wird (Ricken & Reh, 2017). Prüfen wird tendenziell wenig als genuin pädagogische Praktik thematisiert. Dabei ist das Prüfen eng mit dem Lehren und Lernen verbunden. Es ist das Prüfen, das einerseits „das Lehren mit einer Macht ausstattet, die das Lernen nach sich zieht und formatiert“, andererseits dazu beiträgt, das Lehren und Lernen „auf das Prüfbares zu begrenzen und beides gegeneinander zu verkehren“ (Ricken & Reh, 2017, S. 248). Diese „Scharnierfunktion“ des Prüfens ist für Ricken und Reh (2017, S. 248) das wesentliche Indiz dafür, das Prüfen nicht als etwas anderes oder dem pädagogischen Äußerliches zu verstehen, sondern als Element pädagogischen Handelns. Die Hochschullehre ist auf einen solchen Zusammenhang hin noch nicht beobachtet worden, sieht man von Diskussionen zu der auf das Erbringen der Prüfungsleistung fokussierten Diskussion zu ‚Teaching-to-the-Test‘ (siehe Beitrag von Hauser in diesem Heft) einmal ab. Zukünftig könnte untersucht werden, wie Prüfungsformate – mündliche Prüfungen, Hausarbeiten, Klausuren – vorauslaufend die Darstellung und Weitergabe von Wissen in Lehrveranstaltungen prägen. An einem besonderen Fall – sogenannten Prüfungsersatzleistungen – kann jedoch gezeigt werden, wie Prüfungsmodalitäten in Modulen die Seminarpraxis konfigurieren.

In Modulordnungen – nicht nur der Geschichtswissenschaft – kommt häufig vor, dass sich Studierende entscheiden müssen,

„in welchem der einem Modul zugeordneten Seminare sie ihre Prüfungsleistung ablegen (z. B. eine Hausarbeit). Im Umkehrschluss besuchen sie eines der Seminare mit abschließender Prüfungsleistung und eines ohne, wobei für das Bestehen des letzteren immerhin die Bedingung der ‚aktiven Teilnahme‘ erfüllt sein muss (regelmäßige Präsenz, Dokumentation der Pflichtlektüre etc.).“ (Kalthoff & Engert, 2021, S. 251)

Lehrende machen dabei die Erfahrung, dass diese heruntergeschraubten Anforderungen an die Teilnahme zu einer nur unverbindlichen Seminarteilnahme führen können. Gegengesteuert wird vielerorts mit einer Prüfungsvorleistung – einer unbenoteten Mini-Hausarbeit, einem Protokoll o. Ä. – die die Studierenden dazu bringen soll, am Seminar aktiv teilzunehmen. Kalthoff und Engert (2021, S. 251) interpretieren dies als „eine Kompensationswährung, die zur Rückbindung der Aufmerksamkeit der Student/innen an den Seminarverlauf und auch an die Gunst der Bewertenden genutzt wird.“ Von Seiten der Studierenden berichten Kalthoff und Engert (2021) davon, dass vor dem

Hintergrund einer ihnen abgeforderten Entscheidung für den einen oder anderen Prüfenden im Modul Einschätzungen zu der Höhe der Anforderungen der Lehrenden rege ausgetauscht würden. Sie weisen zudem auf den Freiheitsverlust auf Seiten der Studierenden als kontraproduktiv für intrinsische Motivation hin, die sich im Studium entwickeln sollte: „Studierende folgen im Verlauf ihres Studiums damit weniger der Entfaltung ihrer Neigungen, sondern – in Anbetracht des Anziehens der Prüfungszwänge bei gleichzeitiger Zeitverknappung – nachvollziehbarerweise dem Kalkül des Minimum-Input/Maximum-Output“ (Kalthoff & Engert 2021, S. 252). Die Verteilung von Prüfungen und die Prüfungsmodalitäten formatieren also die Lehre und das Studium und somit auch die Art der Auseinandersetzung mit den Inhalten – in Teilen – mit.

Programmatisch haben viele Hochschulen in den letzten 10 Jahren große Anstrengungen unternommen, der Lehre eine gestiegene Bedeutung zuzuweisen (z. B. durch das Ausrufen von Lehrpreisen). Demgegenüber haben Kalthoff und Engert (2021) jedoch in ihrer Studie zu Lehr- und Prüfungsalltag Anzeichen für geradezu einen Rückzug – von Studierenden und Hochschullehrenden – aus der Lehre gefunden. Das heißt, dass die verstärkt eingeforderten Prüfungen und deren immer genauere Regulierung und somit die permanent im Studium gesetzten Leistungsanreize bei Studierenden nicht unbedingt dazu führen, sich die Lehre einzubringen. Hochschullehrende ziehen sich wiederum auf die forschungsaffinen Bereiche in der Lehre zurück. „Während Studierende zunehmend auf ‚Punkte‘ schielen, peilen Lehrende zunehmend ihre Forschung respektive Lehranstrengungen für den wissenschaftlich orientierten Nachwuchs an (etwa forschungsorientierte Master-Programme, Graduiertenkollegs, Forschungskolloquien).“ (ebd., S. 252/253). Gerade am Anfang des Studiums, wo in das Fach eingeführt und erste Auseinandersetzung mit dem Fach initiiert werden sollten, sind Studierende hauptsächlich mit der Bewältigung der hohen Prüfungsdichte beschäftigt, während Hochschullehrende tendenziell versuchen, sich diesem Bereich in der Lehre zu entziehen.

Zur Situation der Prüfung

Wenn Prüfen als soziale, damit kommunikative und interaktive Praxis untersucht wird, werden Elemente von Prüfungen sichtbar ge-

macht, die mehr mit der Situation, denn mit dem Ziel einer Prüfung zusammenhängen. Im Folgenden werden einige Befunde in Bezug auf mündliche und schriftliche Prüfungen aufgerufen, die zeigen, welche Anforderungen und Bewältigungsstrategien unterschiedliche Prüfungsformate mit sich bringen.

In Bezug auf mündliche Prüfungen rekonstruiert Tyagunova (2021, S. 186) eine „Verfahrenskompetenz“ auf Seiten der Geprüften. Über die Inhalte der Prüfung hinaus müssen sie sich als versiert und eloquent auch darin erweisen, mit der im Gesprächsverlauf weitgehend offen gehaltenen evaluativen Rückmeldung zu ihren Antworten umzugehen. Anders gesagt, erhalten die Geprüften im Gespräch wenig Rückmeldung dazu, wie der Prüfende ihre Antworten bewertet, dabei sind auch die Erfolgskriterien für die Prüfenden im Gespräch nicht transparent. Auf Basis der Analyse transkribierter Gesprächsverläufe zeigt Tyagunova (2021) die verschiedenen Gesprächstechniken auf, mit der Geprüfte diese Offenheit zu bearbeiten versuchen: Sie vermeiden Schweigen, sie suchen Übereinstimmungen mit dem Prüfenden herzustellen, sie registrieren die Reaktionen der Prüfenden aufmerksam, sie führen Begeisterung auf. Diese Art der Verfahrenskompetenz ist „im Verfahren selbst situiert“ (ebd., S. 218).

Ein spezifischeres Format, mündliche Prüfungen zu Studienprojekten in Praxisphasen in den lehramtsbezogenen Studiengängen, untersuchen Herzmann und Liegmann (2020). Sie beschreiben an Auszügen aus Prüfungsgesprächen, wie im Gespräch selbst die (hohe) Identifikation der Studierende mit ihrem Studienprojekt zugleich vorausgesetzt und hergestellt wird. So sind auch Praktiken eines „sich Zu-Eigen-Machen“ (ebd., S. 737) mit einer damit einhergehenden Demonstration wissenschaftlicher Expertise ebenso ein Element des Prüfungsgesprächs wie die „Unterwerfung an die Prüfungsförmigkeit der Situation“ (ebd., S. 739). Als Geprüftes wird letztlich eine „Demonstration von Wissenschaftlichkeit“ (Herzmann & Liegmann, 2020, S. 242) im Zusammenhang mit einer „Positionierung der Studierende als Forschende“ (ebd., S. 242) und Bemühungen um einen Wissenstransfer in die Praxis rekonstruiert. Deutlich wird, dass auch die Gegenstände der Prüfungen, hier studentische Forschung als Prüfungsinhalt, das Gesprächsformat mitstrukturiert.

Studentische Praktiken der Vorbereitung auf schriftliche Prüfungen untersuchen Tyagunova und Greiffenhagen (2020) am Beispiel von Facebook-Posts. Als Pendant dazu, dass Prüfende aus zeitlichen und

ökonomischen Gründen die in Klausuren genutzten Prüfungsaufgaben wiederverwenden, sind studentische Praktiken der Weitergabe von Altklausuren und ihrer Aneignung zu beobachten. Aus Dozierenden-sicht ist dies nicht zuletzt deswegen ein oft unerwünschtes Verhalten, weil es Dozierende zu der Entwicklung immer neuer Prüfungsfragen – mit dafür aufzubringender Zeit – zwingt. Im Graubereich gesetzlicher Vorgaben findet sich die studentische Kommunikation über Altklausuren mittlerweile auch in den sozialen Netzwerken. Jenseits der Frage danach, ob dies – im Sinne einer Variante von Teaching-to-the-Test (vgl. Beitrag von Hauser in diesem Heft) als Learning-to-the-Test – eine dem Studium angemessene Praktik ist, zeige sich an der Kommunikation der Studierenden über die (Alt-)Prüfungen auch deren Bemühungen, mit ihren limitierten Ressourcen an Zeit und Anstrengung im Studium zu haushalten (Tyagunova & Greiffenhagen, 2020). Im Zentrum steht dann oft das Bestehen der Prüfung. Vom Prüfenden gestellte Fragen zu kennen, bietet dabei ein hohes Maß an Orientierung für die Studierenden. Dabei erfordert der Umgang mit den Prüfungsfragen keineswegs nur ein Auswendiglernen, sondern auch ein Verstehen und somit auch Auseinandersetzungen mit den Fragen.

Auch zu den Praktiken der Prüfenden an Hochschulen ist noch wenig bekannt. Tsarouha (2016) weist auf Basis von Experteninterviews darauf hin, dass die Prüfenden angeben, das Prüfen meist durch *Learning by Doing* erlernt zu haben. Es ist Teil der wissenschaftlichen Sozialisation, als Beisitzer*in an Prüfungen teilzunehmen, Klausuren zu kontrollieren etc. Zwar gibt es mittlerweile hochschuldidaktische Angebote, aber wenig Auseinandersetzung mit dem Thema Prüfungsgestaltung. Die Effekte davon sieht Tsarouha (2016) über die wissenschaftlichen Disziplinen hinweg auch in einer auch wenig reflektierten Praxis der Notenvergabe. Darüber hinaus entwickeln Prüfende beispielsweise Strategien für den Umgang mit Belastungen, die durch Prüfungen entstehen. Sie sehen sich in ihren Bewertungsaufgaben mitunter auch Unzulänglichkeiten ausgesetzt, etwa wenn Hausarbeiten sprachlich unverständlich oder äußerst fehlerhaft verfasst sind, so dass Verständlichkeit und Lesbarkeit schwer beeinträchtigt sind. In Reaktion darauf entwickeln Prüfende in dem von Kalthoff und Engert (2021, S. 260) untersuchten universitären Fach Geschichtswissenschaft „Devianzkategorien mit unterscheidbarem Ausmaß“, die legitimierend für die „Zurückweisung einer eingehenderen Befassung mit studentischen

Texten“ und zugleich für „bestimmte Mindeststandards“ im eigenen Fach (hier Geschichtswissenschaft) sorgend eingesetzt werden.

Fazit

Was lässt sich mit diesen Innensichten zum Prüfen als soziale Praxis und die damit gewonnenen Einsichten anfangen? Sie könnten zu Reflexionen führen zu der Frage, wie Prüfungen den Alltag von Lehre und Studium mitstrukturieren. Diskutiert werden könnte, welche dys/funktionalen Effekte von Prüfungen für den Alltag von Lehre und Studium ausgehen. Dabei würden einmal nicht die Kompetenzen von Individuen ins Zentrum gerückt werden, also nicht gefragt werden, wie Lehrende Prüfungen noch besser standardisieren und Bewertungen objektivieren sollten bzw. Geprüfte ihre kognitive Leistungsfähigkeit steigern könnten. Vielmehr wäre zu diskutieren, wie Prüfungen den Alltag im Studium und in der Lehre bestimmen und wie sie welchen Umgang mit Wissen im Studium und in der Lehre prägen. Es würde darum gehen, den Zusammenhang der in das Studium eingelassenen kleinschrittigen Dauerbewertungen und den sich im Zusammenhang mit ihnen entwickelnden Umgangsweisen von Hochschullehrenden und Studierenden zu verstehen. Dabei wäre auch über unerwünschte Effekte von Prüfungen zu reflektieren, ohne einseitig auf Seiten von Individuen nach Ursachen zu suchen. Voraussetzung wäre, mit der Reflexion der alltäglichen Praxis des Prüfens nicht nur auf Effizienzsteigerung zu zielen, sondern nach dem Einfluss der Praxis des Prüfens in einem weiten Sinne auf die Qualität von Lehre und Studium zu fragen.

Literatur

- Herzmann, P. & Liegmann, A. (2020). Mündliche Prüfungen im Kontext des Forschenden Lernens. (Re-)Adressierungen als Inszenierung studentischer Expertise. *Zeitschrift für Pädagogik*, 5(66), 727-745.
- Kalthoff, H. & Engert, K. (2021). Universitäre Bewertung. Die Geschichtswissenschaft und die Einsätze ihrer Mitglieder. In F. Meier & T. Peetz (Hrsg.), *Organisation und Bewertung* (S.239-271). Wiesbaden: Springer VS.
- Ricken, N. & Reh, S. (2017). Prüfungen – Systematische Perspektiven der Geschichte einer pädagogischen Praxis. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 63(3), 247-259.

- Tsarouha, E. (2016). *Prüfungspraktiken an deutschen Hochschulen. Eine empirische Studie zu systematischen Einflussgrößen auf die Notengebung in Abschlussprüfungen*. Wiesbaden: Springer VS.
- Tyagunova, T. (2021). Prüfungskompetenz. Interaktive Steuerung von Wissensdarstellungen in mündlichen Universitätsprüfungen. *Sozialer Sinn*, 22(1), 185-221.
- Tyagunova, T. & Greiffenhagen, C. (2021). ‚Learning for the test‘: A study of students’ sharing and discussing past multiple choice questions on Facebook. *Ethnographic Studies*, 18, 43-74. <https://doi.org/10.5281/zenodo.5805469>

*Kerstin Rabenstein, Dr., Professorin
am Institut für Erziehungswissenschaft,
Georg-August-Universität Göttingen.*

*Arbeitsschwerpunkte:
Transformation pädagogischer Ordnungen,
qualitative Schul- und Unterrichtsforschung*



kerstin.rabenstein@sowi.uni-goettingen.de

05

Michaela Kaiser

Mythos Eignungsprüfung.
Kunst – Leistung –
hochschulische Prüfungspraxis

Die Aufnahme des Studiums des künstlerischen Lehramts ist in Deutschland an spezifische Voraussetzungen geknüpft: Unabhängig von der Hochschulreife müssen Studienbewerber*innen am jeweiligen Studienstandort eine kunstpraktische Eignungsprüfung aus Mappenprüfung, praktischer Prüfung und Gespräch absolvieren. Ihr kommt als Zugangsvoraussetzung eine tragende Rolle zu, doch ein Diskurs über die Feststellung der künstlerischen Eignung wird kaum geführt. Mit diesem Beitrag wird von daher das Ziel verfolgt, diesen – angeregt durch einen Auszug eines Fachgesprächs – mittels eines Problemaufrisses anzustoßen und auf der Grundlage der Sichtung exemplarisch ausgewählter Studienordnungen, Informationen zur Praxis der Eignungsprüfung an deutschen Hochschulen bereitzustellen.

Anforderungsreiche Prüfungskultur

Eignungsprüfungen im Fach Kunst bestehen bereits seit mehreren Jahrhunderten – um es genau zu nehmen wurde schon im 16. Jh. an der u. a. von Vasari mitbegründeten Kunstakademie in Florenz die künstlerische Eignung der zu dieser Zeit noch ausschließlich männlichen Bewerber gesichtet. Die Gründung von Kunstakademien entwickelte sich nach und nach und damit etablierte sich ebenfalls die Praxis der Eignungsfeststellung. Doch nicht nur auf Seiten von Studienbewerber*innen, sondern auch auf Seiten der hochschulischen Akteur*innen stellt die Eignungsprüfung eine Herausforderung dar, in der sich die zentralen Spezifika des Faches hinsichtlich des Umgangs mit Leistung abbilden: neben dem Für und Wider des Instruments selbst erweist sich die Herausforderung der *Standardisierung des Künstlerischen* und dessen *Offenheit, Unbestimmtheit und Ereignishaftigkeit* als Spannungsfeld, mit dem Akteur*innen an Hochschulen konfrontiert sind. Gängige Kritik ist, dass die hierarchisierende und selektierende Funktion der Leistungsmessung Prozesse der (künstlerischen) Normierung abverlangt, wobei es in der Kunst gerade um eine Auflösung dieser dialektischen Positionen geht und vielmehr die Kontingenz des Künstlerischen in der Beschreibung *guter* Kunst hervorgehoben wird (u. a. Engel, 2020). Auch die unterschiedlichen Anforderungen an die Prüfungspraxis an den Hochschulen tragen zu Konfusion bei. Vor allem eine merkbare Uneinheitlichkeit in Bezug auf hochschulische Erwartungen verunsichert: So bleiben die Anforderungen an vielen Stellen unscharf und

die standortspezifischen Umsetzungen der Eignungsprüfung unterscheiden sich teils stark voneinander, wie exemplarisch in Tabelle 1 abgebildet. Zwar tendieren die Institute hinsichtlich der qualitativen Anforderungen mehrheitlich zu den traditionellen Techniken wie Malerei, Zeichnung, Druckgrafik, Fotografie und Skulptur, dennoch sind hinsichtlich der Zusammenstellung der Arbeiten Unterschiede vorhanden: Im Hinblick auf die Mappenprüfung werden an einigen Standorten mehr, an anderen weniger eingereichte Werke eingefordert; die praktische Prüfung stellen die einen unter ein Thema und wieder andere legen deutlich mehr Wert auf spezifische technische Annäherungen und entfernen sich vom traditionellen Kanon künstlerischer Techniken. Die Beurteilung der künstlerischen Eignung scheint insgesamt wenig transparent: In sämtlichen Prüfungsordnungen fehlen Angaben zu den inhaltlichen Erwartungen an Mappe, praktische Prüfung und mündliches Gespräch bzw. werden von der Aussage verdeckt, dass über die eingereichten Arbeiten ein Eindruck der eigenen Person vermittelt werden soll. Der zeitliche Umfang der Prüfung variiert zwischen vollen zwei Tagen wie an der Akademie in München, in denen Prüfungen und Gespräche stattfinden, und drei Stunden wie am Standort Osnabrück. An einigen Standorten kann die Allgemeine Hochschulreife durch eine herausragende künstlerische Eignung ersetzt werden, so dass auch eine fachgebundene Hochschulreife das Studium bei bestandener Eignung ermöglichen kann.

Tab. 1 Übersicht über die Prüfungsanforderungen exemplarisch ausgewählter Hochschulstandorte

	Frankfurt	Osnabrück	Köln	München	Potsdam
Organisationsform	1. und 2. Staatsprüfung	Bachelor/ Master	Bachelor/ Master	1. und 2. Staats-examen	Bachelor/ Master
Vorgaben der Ministerien	laut Kultus- und Wissenschaftsministerium Auslegungssache der Hochschulen			laut Kultus- und Wissenschaftsministerium Auslegungssache der Hochschulen (Verweis auf Kunsthochschulengesetz)	

Weitere Voraussetzungen	Hochschulreife/NC	Hochschulreife/NC	Hochschulreife/NC	Hochschulreife	Abitur/NC
Mappe					
Umfang	10 selbstständig angefertigte Arbeiten	20 selbstständig angefertigte Arbeiten	20 selbstständig angefertigte Arbeiten	25-30 selbstständig angefertigte Arbeiten	20 selbstständig angefertigte Arbeiten
Thema	eigene Themenwahl	eigene Themenwahl	eigene Themenwahl	eigene Themenwahl	eigene Themenwahl
Bewertungskriterien	eigener künstlerischer Stil soll Werken ablesbar sein	eigener künstlerischer Stil soll Werken ablesbar sein	Mappe soll einen Eindruck der Person wiedergeben	eigener künstlerischer Stil soll Werken ablesbar sein	eigener künstlerischer Stil soll Werken ablesbar sein
Prüfung					
Thema	vorgegeben	vorgegeben, keine Wahlmöglichkeit	vorgegeben, keine Wahlmöglichkeit	vorgegeben, 2 Auswahlmöglichkeiten	vorgegeben, 3 Auswahlmöglichkeiten
Technik	-	Malerei, Zeichnung, Plastik/Skulptur	Video von 3 min Länge	-	Malerei, Zeichnung, Plastik/Skulptur, Fotografie
Dauer	-	3 Stunden	-	2 Tage	2 Tage
Bewertungskriterien	-	-	-	-	-
Gespräch					
Inhalt	-	-	entfällt	-	-
Dauer	-	20 min	entfällt	-	20 min

Um die gegenseitige Anerkennung der Eignungsprüfung bei den Hochschulen ist es ungewiss bestellt: ob Studierende des Lehramts Primarstufe, welche an der Universität Rostock ohne eine Eignungsprüfung zum Studium zugelassen wurden, an die Kunstakademie München wechseln können, die zur Aufnahme eine umfangreiche Mappe, eine mehrtägige Prüfung sowie ein Prüfungsgespräch einfordert, obliegt – einzelfallgeprüft – der Hochschule selbst.

Die Uneinheitlichkeit der Prüfungskulturen spiegelt die unterschiedlichen und teils widersprüchlichen regulativen Verordnungen auf den

verschiedenen Ebenen des Systems Hochschule wider. Die Eignungsprüfung für das Fach Kunst ist eingebettet in ein System, in welchem die Anforderungen an die Prüfungsleistung auf unterschiedlichen Vollzugsebenen von Bildungspolitik, Bildungsverwaltung, Hochschulpolitik und Hochschulverwaltung sowie den Akteur*innen an den Hochschulen selbst und den Studienbewerber*innen übersetzt werden. Während die *Systemlogik* Akteur*innen an Hochschulen zur Normierung der künstlerischen Leistung und zur Präskription der hier ablesbaren künstlerischen Entwicklung auffordert, geht es der *Binnenlogik* der künstlerischen Eignung nach daneben um die Kontingenz des Künstlerischen, die der Normierung künstlerischer Leistung diametral gegenübersteht sowie um die Unvorhersagbarkeit der künstlerischen Entwicklung im Rahmen des Studiums (u. a. Engel, 2020; Kaiser & Brenne, 2020).

Spannungsfelder der Prüfungspraxis

Solche Handlungsdilemmata bilden sich auch im folgenden Fachgespräch ab:

„Wie sollen die eingereichten künstlerischen Mappen und die Prüfungsleistung beurteilt werden, welche Kriterien werden zugrunde gelegt?“ mehren sich die Fragen der Mitarbeitenden. In den Vorgaben des Instituts finden sich die entscheidenden Kriterien: „originell und handwerklich anspruchsvoll sollen die Arbeiten gestaltet sein, sie sollen in ihrer Gesamtheit eine künstlerische Positionierung veranschaulichen und insgesamt von Stringenz zeugen“, zitiert eine Mitarbeitende die Prüfungsordnung. Schnell sind sich die Kolleg*innen einig, dass hier auf hochschulischer Ebene der Versuch unternommen wurde, die Empfehlungen der Kultusminister umzusetzen: „am grünen Tisch seien die Kriterien entworfen worden, doch erfahrene Kolleg*innen müssten nur einen kurzen Blick in eine Mappe werfen und wüssten sofort, ob sich eine Person für das Studium eigne, dazu bedarf es keines Kriterienkatalogs“. Andere wiederum sind dankbar für möglichst klar formulierte Bewertungskriterien, die persönliche Unsicherheiten auffangen und hierüber ein gerechtes Urteil ermöglichen. „Doch was impliziert also das Kriterium der Originalität? Und was ist eine ‚gute‘ künstlerische Position?“ lässt eine Mitarbeiterin nicht locker, „kann man das überhaupt unter dem Einfluss des eigenen Geschmacks gerecht beurteilen? Dürfen auch Arbeiten aus dem Schulunterricht enthalten sein, oder büßt die Mappe dadurch an Authentizität

und Originalität ein? Dem folgt die nächste Frage einer anderen Mitarbeiterin: „Wie ist mit den eingereichten Mappen umzugehen, die aufgrund der Pandemiesituation digital an zahlreiche Studienstandorte versendet wurden?“ Muss hier bei der Auswahl der Bewerber*innen und der Zulassungen eine Ausfallquote einkalkuliert werden? Es wird dann weiter eingewendet, dass es möglicherweise auch die aktuellen Notstände in Bezug auf ausgebildete Kunstlehrkräfte in der Grundschule zu berücksichtigen gelte: „Sollte diesbezüglich einer großzügigeren Auslegung der Kriterien gefolgt werden? Oder spielen solche organisationalen Entwicklungen bei der Begutachtung der eingereichten Arbeiten keine Rolle?“ fragen sich zwei Mitarbeitende. „Wozu eigentlich eine Eignungsprüfung, wenn nicht ausschließlich die künstlerische Eignung im Fokus der Prüfung steht, sondern auf der Hinterbühne andere Dynamiken ihren Einfluss nehmen“, so der Einwand: „Lasst es uns doch einfach so machen wie Beuys!“ (Kaiser, 2021, Z. 105-139)

Als Professor an der Kunstakademie Düsseldorf betreute Beuys eine große Zahl Studierender. Hier vertrat er die Meinung, dass jede*r, der*die Kunst studieren wolle, dies auch können solle. Das Bewerbungsverfahren aus Mappenprüfung und künstlerischer Eignungsfeststellung, wie es bis heute üblich ist, lehnte er ab und nahm stattdessen abgelehnte Bewerber*innen auf. Er wurde verwarnt, besetzte unter Bezugnahme auf Freiheit in Forschung und Lehre aber gemeinsam mit abgewiesenen Studierenden die Kunstakademie und wurde schließlich vom Wissenschaftsminister Johannes Rau wegen Hausfriedensbruchs entlassen.

Die auf regulativer Ebene erlassene Verpflichtung zur Überprüfung und Bewertung der künstlerischen Eignung übersetzt sich auf die Ebene der Akteur*innen an Hochschulen und stattet sie mit Spannungsfeldern aus:

Normierung – Kontingenz

Für hochschulische Akteur*innen bedeutet es eine große Herausforderung, im *Spannungsfeld von Normierung und Kontingenz* zu agieren. Wie mit der Rekonstruktion des Fachgesprächs nachgezeichnet, sind Akteur*innen an Hochschulen mit widersprüchlichen Zielerfordernissen – der Normierung offener und unbestimmter künstlerischer Prozesse – konfrontiert. Aus diesen ergeben sich Handlungsdilemmata, wenn einerseits die Originalität und die individuelle künstlerische Positionierung von Studienbewerber*innen zu bewerten ist und

andererseits zur Beurteilung einer geforderten nicht-normativen Ästhetik vergleichende Kriterien herangezogen werden. Auf der einen Seite sollen Originalität, Authentizität und eine eigene künstlerische Positionierung anerkannt und gewürdigt werden, auf der anderen Seite die individuellen künstlerischen Positionierungen sachlich miteinander verglichen werden. Hochschuleitig wird daher in der Regel im Hinblick auf qualitative Aspekte keine Einschränkung bei den eingereichten Arbeiten vorgenommen und teilweise explizit dazu aufgefordert, eigenständige Arbeiten vorzulegen, die sich hinsichtlich der Gestaltung von jenen des Schulunterrichts unterscheiden. Die Mappe soll hierdurch *unbeeinflusster und authentischer* die künstlerische Eignung abbilden. Dieser institutionellen Perspektive steht jedoch gegenüber, dass die Eignungsprüfung und die Erstellung der Mappe oftmals ein höchst diskursiver sozialer Prozess sind: „jene, die sich bei der Mappengestaltung ausschließlich auf ihr eigenes Relevanzsystem verlassen, (sind) eher schlecht beraten“ (Rothmüller, 2012, S. 99). Schon an dieser Stelle zeigt sich, dass jene Bewerber*innen, die solche institutionellen Bewertungspraktiken kennen, einen Vorteil haben (vgl. ebd.), was dem Anspruch entgegensteht, dass Originalität und künstlerische Positionierung am gleichen sachlichen Bezugspunkt objektiv gemessen werden.

Zwar zeigt sich im Rückblick in die Fachgeschichte eine regelmäßige Adaption unterschiedlicher Bezugsnormen, die den Umgang mit künstlerischer Leistung prägen (Otto, 1972; Peez, 2018, 2013; Seydel, 2003) und Eingang in die hochschulischen Prüfungsordnungen gefunden haben, doch ist bei der Beurteilung handwerklicher Fähigkeiten, der Originalität und Authentizität oder der künstlerischen Positionierung stets die subjektive Evidenz von Beurteilenden, ihr individuelles ästhetisches Urteil, implizit hinterlegt (Peez, 2004). So genügt einigen ein kurzer Blick in die Mappe und sie wissen *sofort, ob sich der*die Bewerber*in für das Studium eigne*. Zwar ist diesbezüglich im Sinne Kants (1790) die Legitimierung über ein intersubjektives ästhetisches Empfinden angesprochen, gerät jedoch in Verflechtung mit den kulturtheoretischen Ausgangspunkten Bourdieus (1987) und den hiermit in Verbindung stehenden habituellen Orientierungen von Prüfenden in die Kritik (Peez, 2004, 2013). Das ästhetische Urteil erweist sich somit als verhandelbare Konstruktion, das sich sowohl historisch – denkt man an die aktuelle Pandemiesituation oder den Lehrkräftemangel an Grundschulen – als auch kulturell, im Hinblick auf die Einschätzung der

künstlerischen Positionierung, unterscheidet (Kaiser & Brenne, 2021). So absolvieren Bewerber*innen mit hohem kulturellen und sozialen Kapital häufiger erfolgreich als andere die Aufnahmeprüfung für das Fach Kunst, da sie mit den institutionellen und habituellen Praktiken stärker vertraut sind und sich im Hinblick auf die Gestaltung der künstlerischen Werke auf unterschiedliche kulturelle Bezugssysteme beziehen können (Rothmüller, 2012). Vor diesem Hintergrund werden für Studienbewerber*innen über den *Zwang* zur Normierung der künstlerischen Leistung und der gleichzeitigen Betonung derer Kontingenz die Zugänge zur weiteren Entfaltung der künstlerischen Position ermöglicht oder auch verhindert.

Prognosen – Unvorhersagbarkeit

Im Zuge der Feststellung der künstlerischen Eignung wird erwartet, dass Akteur*innen an Hochschulen eine Prognose über die Eignung für das Lehramtsstudium Kunst abgeben. Doch bleibt dabei unberücksichtigt, dass im Zuge der Feststellung der künstlerischen Eignung immer nur Ausschnitte erfasst werden können, insofern die hochschulische Entwicklung einer eigenen künstlerischen Position als außerordentlich komplexer Prozess und dynamischer Vorgang zu bezeichnen ist. Erschwerend kommt hinzu, dass bei der Feststellung der künstlerischen Eignung unberücksichtigt bleibt, dass vielfach der Lehrberuf die berufliche Zielperspektive vorgibt, eine Eignung explizit für die pädagogische Tätigkeit aber nicht geprüft wird. So bleibt die Frage nach den überfachlichen Kompetenzen in der Feststellung der künstlerischen Eignung offen (Peez, 2018); auch wird neben der kunstpraktischen Eignung eine kunsttheoretische Eignung nicht geprüft und es stellt sich die Frage, ob solche Fähigkeiten für den Kunstlehrkräfteberuf weniger wichtig sind. Das Fach Kunst testet die Studienbewerber*innen so gesehen nur in einem Ausschnitt ihrer Eignung, die sie bereits vor dem Studium mitbringen, auf die allein sie aber in der späteren Berufslaufbahn nicht setzen können.

Offene Fragen

Von Studieninteressierten für das Fach Kunst wird neben der Hochschulreife – ob allgemein oder fachgebunden – auch der Nachweis der besonderen Eignung vorausgesetzt, der zertifiziert, dass sie die fach-

lichen Grundvoraussetzungen mitbringen, um ein Studium in diesem Fach bestreiten zu können. Hier schließen die offenen Fragen an, ob

- die Hochschulreife die für das Studiums des Fachs notwendigen Qualifikationen nicht abbilden kann,
- die künstlerische Eignung als ein angeborenes Talent verhandelt wird, welches nicht vermittelt und somit
- grundsätzlich nicht im Studium entwickelt werden kann?

Für das Studium des Fachs Kunst ist diesen Fragen bisher weder theoretisch noch empirisch nachgegangen worden, so dass fundierte Antworten ausstehen. Und so bleiben viele Fragen offen – auch, da neben der kunstpraktischen die kunstwissenschaftliche und kunstpädagogische Eignung nicht geprüft wird und nach wie vor unter Berücksichtigung sozialer Ungleichheitsverhältnisse Homogenisierungsprozesse stattfinden.

Literatur

- Bourdieu, P. (1987). *Die feinen Unterschiede*. Frankfurt am Main: Suhrkamp.
- Engel, B. (2020). Unbestimmtheit als (kunst-)didaktisches Movens in professionsbezogenen Bildungsprozessen. In B. Engel & K. Böhme (Hrsg.), *Didaktische Logiken des Unbestimmten* (S. 58-85). München: kopaed.
- Kaiser, M. (2021). Ethnografisches Protokoll eines Fachgesprächs im Rahmen des Projekts Leitbilder künstlerischer Leistung. [unveröffentlichtes Protokoll]
- Kaiser, M. & Brenne, A. (2020). Kontingenzbearbeitung – Regulativ inklusiver kunstpädagogischer Professionalisierung. *Qfl*, 2(1). <https://doi.org/10.21248/Qfl.22>
- Otto, G. (1972). Anmerkungen zur aktuellen Problematik von Leistung und Zensur – nicht nur im Kunstunterricht. *K+U*, 16, 43-45.
- Peez, G. (2004). Ästhetisches Urteil und Evidenzurteil. *K+U*, 287, 38.
- Peez, G. (2013). Beurteilen als kulturelle Sinnpraxis. Ästhetische Urteilsbildung als Kompetenz im Kontext. In: S. Engels, R. Preuss & A. Schnurr (Hrsg.), *Feldvermessung Kunstdidaktik*. (S. 79-91). München: kopaed.
- Peez, G. (2018). *Einführung in die Kunstpädagogik* (5. Aufl.). Stuttgart: Kohlhammer.
- Rothmüller, B. (2012). *Soziale Barrieren beim Zugang zu einem künstlerischen Studium*. Verfügbar unter <https://orbi.uni.lu/bitstream/10993/3361/1/Rothmueller%20Soziale%20Barrieren%20Proof%20Beltz.pdf> [28.08.2021].
- Seydel, F. (2003). Beurteilen lernen. Diskussionsbeitrag zu einem leidigen Thema: Zensuren im Kunstunterricht. *BDK-Mitteilungen*, 1, 34-35.

Michaela Kaiser, Dr., akademische Mitarbeiterin
am Department für Ästhetische Bildung
an der Universität Potsdam.
Arbeitsschwerpunkte:
Inklusion und (Leistungs-)Differenz
in kunstpädagogischen Kontexten



michaela.kaiser@uni-potsdam.de

06

*Annette Bauer-Klebl
und Charlotte Nüesch*

Prüfungsevaluationen und damit
verbundene Feedbackgespräche:
Wie sie zur Hochschulentwicklung
beitragen können

Einleitung

Obwohl den Prüfungen ein hoher Einfluss auf das Lehr-Lerngeschehen attestiert wird, ist Prenzel (2015) der Ansicht, dass die Steuerungsmöglichkeiten von Prüfungen auf das Lehren und Lernen an den Hochschulen bislang noch weitgehend ungenutzt geblieben seien. Frölich-Steffen (2019) vermutet, dass in der Vergangenheit die Potenziale zur Gewährleistung einer hohen Qualität der Hochschullehre noch zu wenig ausgeschöpft worden seien. Vor diesem Hintergrund stellt sich die Frage, wie diese Potenziale besser genutzt werden könnten. Ein möglicher Ansatz zur Förderung des Constructive Alignment innerhalb von Modulen und innerhalb eines Studiengangs können systematische Evaluationen summativer Lernerfolgskontrollen sein. Dieser Beitrag beschreibt die Erfahrungen einer Schweizer Fachhochschule mit einem solchen systematischen Ansatz, der die Qualität von Studiengängen eines Departements nachhaltig sichern und weiterentwickeln will. Hierzu werden die zentralen Elemente des Evaluationskonzepts eingebettet in die relevante Theorie vorgestellt. Anschliessend werden die damit im Verlaufe der vergangenen rund zehn Jahre gemachten Erfahrungen kritisch reflektiert und Potenziale für die Weiterentwicklung identifiziert.

Evaluationskonzept

Das Konzept zur Evaluation summativer Lernerfolgskontrollen ist eingebettet in ein ganzheitliches Qualitätsverständnis. Dieses verfolgt das Ziel einer umfassenden Weiterentwicklung der Hochschullehre und orientiert sich an einem Input-Prozess-Output-Modell (Hoy & Miskel, 2013). Das Evaluationskonzept setzt bei der Prozessperspektive an und fokussiert das konkrete Verhalten der Dozierenden bei der Konstruktion von Leistungsnachweisen. Inputseitig sind zwei Gestaltungselemente bedeutsam: (1) Das Bildungs- und Lehr-Lernverständnis, das die zentralen Eckpfeiler für die Gestaltung von Lehr-Lernprozessen schriftlich festhält, dient den Hochschulangehörigen als Leitschnur bei der Gestaltung des Kontaktunterrichts, des begleiteten Selbststudiums und der Leistungsnachweise. (2) Neue Dozierende ohne ausreichende didaktische Vorkenntnisse besuchen obligatorisch einen CAS-Studiengang in Hochschuldidaktik, bei dem die valide und reliable

Gestaltung von Leistungsnachweisen einen zentralen Bestandteil darstellt.

Die Entwicklung des Evaluationskonzepts erfolgte mit Hilfe eines morphologischen Kastens, der anhand relevanter Kriterien (vgl. Gomez, Kobler & Nüesch, 2011) die grundsätzlich möglichen Ausprägungen und das gewählte Konzept (grau hinterlegte Felder) zeigt.

Tab. 1 Morphologischer Kasten zur Ausgestaltung des Evaluationskonzepts

Kriterien	Ausprägungen				
Zeitpunkt	Vor der Durchführung		Während der Durchführung		Nach der Durchführung
Anlass	Aufgrund von Beanstandungen		Zufällig		Gemäss Turnus
Zielgruppe	Bachelor (Grundstudium)		Bachelor (Hauptstudium)		Master
Prüfungsform	Schriftliche Prüfungen		Seminararbeiten, Projekte		Mündliche Prüfungsleistungen
Prüfungsphasen	Konstruktion		Durchführung		Auswertung
Evaluierende Personen	Selbstevaluation durch Dozierende (D)	Studierende (S)	Interne Experten (E)	Externe Experten (E)	Gemischte Teams (D, S)
Evaluationsverfahren	Dokumentenanalyse		Statistiken	Befragung	Gespräch
Instrumente	Checkliste für Dozierende		Fragebogen		Liste mit Evaluationskriterien und Erläuterungen

Aus der Tabelle 1 wird deutlich, dass ein breit abgestütztes, formatives Evaluationsmodell gewählt wurde. Gegenstand der Evaluation sind die Module aller angebotener Studiengänge in einem Departement auf allen Stufen, wobei sowohl schriftliche und mündliche Prüfungsleistungen als auch Seminararbeiten und Projekte berücksichtigt werden. Die Evaluation findet vor der Durchführung der Prüfungsleistung statt, so dass noch gezielte Anpassungen möglich sind. Der Schwerpunkt der Evaluation liegt deshalb bei der Konstruktion der Lernerfolgskontrollen und umfasst die Analyse verschiedener relevanter Dokumente (Abgangskompetenzen aus den Modulbeschreibungen, geplante

Prüfungsaufgaben bzw. Projektauftrag, erwartete Lösungsskizze mit Punktezuteilungen bzw. Beurteilungsraster). Für einen Zyklus von zwei bis drei Jahren bestimmen jeweils die Studiengangleitungen in Abstimmung mit den Expert*innen des Didaktikzentrums die zu evaluierenden Module. Ziel ist es, über einen bestimmten Zeitraum hinweg alle Dozierenden des Departements mit Modulverantwortung mindestens einmal in die Evaluation einzubeziehen, so dass die Qualität der Studiengänge breit und nachhaltig weiterentwickelt und die Dozierenden in ihrer Kompetenzentwicklung gefördert werden können. Die hier beschriebene Evaluation ist eine Kombination von Selbstreflexion durch die Prüfenden und Fremdeinschätzung durch interne Expert*innen des Didaktikzentrums. Die Evaluation erfolgt dabei anhand von Beurteilungskriterien, die sich am Grundsatz des Constructive Alignment (Biggs, 1996; Biggs & Tang, 2011) und an den Gütekriterien von Prüfungen (Frölich-Steffen & den Ouden, 2019; Metzger & Nüesch, 2004; Schaper, 2021) orientieren. Die Beurteilungskriterien werden den Prüfenden zur Selbstreflexion in Form von Leitfragen zur Verfügung gestellt (vgl. Tab. 2). Die Ergebnisse der vorgenommenen Beurteilung (Selbst- und Fremdeinschätzung) werden anschliessend in einem Gespräch verglichen, unterschiedliche Sichtweisen und alternative Möglichkeiten diskutiert und gemeinsam Folgerungen für die Gestaltung des Leistungsnachweises abgeleitet. Mögliche Massnahmen aufgrund des Gesprächs reichen von geringfügigen Modifizierungen von Aufgaben für die kurz bevorstehende Prüfung über Anpassungen der Modulbeschreibungen (z. B. taxonomisch höher angesiedelte Abgangskompetenzen) hin zu einer veränderten Gestaltung des künftigen Unterrichts (z. B. ausgeprägtere Kompetenzorientierung, vermehrte Fallstudienarbeit) bis zu einer völligen Neugestaltung der Prüfungsform (z. B. open book-Prüfungen).

Die Evaluationsergebnisse werden den Studiengangleitungen, der Departementsleitung und den Vorgesetzten der evaluierten Dozierenden kommuniziert. Damit wird sichergestellt, dass die Qualitätsentwicklung ins jährliche Mitarbeitendengespräch Eingang findet und bei Bedarf geeignete Entwicklungsmassnahmen (z. B. Weiterbildungskurse, Coachings) vorgesehen werden können.

Tab. 2 Leitfragen zur Selbstreflexion: Schriftliche Prüfung

Validität	<ol style="list-style-type: none"> 1. Habe ich mir einen Überblick über die angestrebten Abgangskompetenzen in der Modulbeschreibung verschafft und diese interpretiert? Inwiefern entsprechen diese dem ausgewiesenen Modulniveau? 2. Streuen die Prüfungsaufgaben angemessen über die Sachgebiete und Themen des Prüfungsgebietes? 3. Ist das Niveau der Prüfungsaufgaben hochschulgerecht, d. h. sind die Prüfungsaufgaben kompetenz- und transferorientiert und dem ausgewiesenen Modulniveau entsprechend formuliert? 4. Sind das gewählte Aufgabenformat und der in der Prüfungsaufgabe geforderte Denkprozess geeignet, um die angestrebte(n) Abgangskompetenz(en) zu überprüfen?
Durchführungsobjektivität	<ol style="list-style-type: none"> 5. Berücksichtigen die Prüfungsaufgaben die Lernbedingungen der Studierenden? 6. Sind die Fragen bzw. Aufträge an die Studierenden eindeutig und präzise formuliert? 7. Habe ich alle zur Lösung der Aufgaben notwendigen Informationsmaterialien zusammengestellt? 8. Ist die Prüfung als Ganzes sinnvoll aufgebaut? 9. Reicht die zur Verfügung gestellte Zeit aus, um die Prüfung zu lösen? 10. Ist die formale Gestaltung der Prüfung angemessen? 11. Sind den Studierenden im Vorfeld der Prüfung die Prüfungsanforderungen transparent und zugänglich?
Auswertungsobjektivität	<ol style="list-style-type: none"> 12. Bestehen im Voraus klare Erwartungen (Kriterien), wie die Aufgaben von den Studierenden beantwortet bzw. gelöst werden sollen (z. B. Lösungsskizze, Einschätzskala)? 13. Besteht für die Bewertung der Aufgaben ein klares, im Voraus definiertes Verfahren (z. B. Regeln für die Punktezuteilung)? 14. Steht die Gewichtung der Aufgaben in einem sinnvollen Verhältnis zum Schwierigkeitsgrad und zur benötigten Zeit?

Zentrale Erkenntnisse der Evaluation und Weiterentwicklungspotenziale

In rund zehn Jahren konnten vielfältige Erkenntnisse mit dem beschriebenen Evaluationskonzept gesammelt werden, die das Ableiten von Folgerungen in Bezug auf die Weiterentwicklung von Studiengängen und die Kompetenzentwicklung der Dozierenden erlauben. Auf *curricularer Ebene* ergab die Evaluation, dass die Abgangskompetenzen in den Modulbeschreibungen häufig angemessen anspruchsvoll formuliert waren. Allerdings zeigten sich Abgrenzungsproblema-

tiken bei aufeinander aufbauenden Modulen – dies sowohl innerhalb der Bachelorstudiengänge als auch an der Schnittstelle zu den darauf aufbauenden Masterstudiengängen. So wurden Theorien und Modelle wiederholt vermittelt, ohne dass dies den Modulverantwortlichen bewusst gewesen wäre. Zudem gab es bezüglich der Modulziele vor allem bei weiterführenden Modulen Niveauschwierigkeiten. So wurde immer wieder zu wenig berücksichtigt, dass das Ausmass an Problemlösung und Transferorientierung deutlich höher sein sollte als in Grundlagenmodulen. Die Studiengangleitungen sollten deshalb die Abstimmung der Module innerhalb des Studiengangs und bei aufeinander aufbauenden Studiengängen als permanenten Prozess der Curriculumsentwicklung im Blick haben und die Reflexion darüber regelmässig in geeigneten Austauschgefässen einfordern.

In Bezug auf die *Erfüllung der Gütekriterien der evaluierten Leistungsnachweise* zeigte sich ein differenziertes Bild. Die *Themengültigkeit* war bei allen Leistungsnachweisen mindestens weitgehend erfüllt, so dass diesbezüglich nur geringfügige Anpassungen notwendig waren. In Bezug auf die *Durchführungs- und Auswertungsobjektivität* wurden unterschiedliche Mängel festgestellt: Ungenauigkeiten bei der Formulierung der Aufgabenstellungen und in den Musterlösungen; zu tiefe Punktezuteilung bei anspruchsvollen Aufgabenstellungen; heikler Umgang mit Ratewahrscheinlichkeit bei Auswahlaufgaben (z. B. intensiver Einsatz von Richtig-Falsch-Aufgaben, MC-Aufgaben, die den Qualitätsanforderungen in Bezug auf die Distraktoren nicht genügen); nicht trennscharfe und ausreichend detaillierte Ausformulierung von geeigneten Kriterien bei Beurteilungsrastern. Teils grosses Entwicklungspotenzial wurde bei der *Konstruktvalidität* identifiziert. Gründe waren meist eine fehlende bzw. für das Niveau der Lernziele ungenügende Transferorientierung, ein fehlender Kontextbezug der Aufgabenstellungen (z. B. unechte Situationsaufgaben, die zur Lösung die Berücksichtigung des Kontexts nicht erfordern; unangemessene Situationskontexte) sowie zu wenig lernblockübergreifende Prüfungsaufgaben, die auf Vernetzung zielen. Es zeigte sich zudem, dass die formulierten Modulziele bei der Gestaltung der Prüfung teils eine untergeordnete Rolle spielten und noch häufig in Inhaltsdimensionen gedacht und geplant wurde. Gerade die Ergebnisse zur Validität decken sich mit Ergebnissen aus den wenigen Studien, die es zur Prüfungspraxis an deutschsprachigen Hochschulen gibt (Schindler, 2015). So konnte etwa Stefancia (2013) bei einem Vergleich zwischen Modulschlussprüfungen und

Modulbeschreibungen zeigen, dass Dozierende Mühe haben, Modulziele mit einem höheren Anspruchsniveau angemessen in Prüfungen abzubilden. Während die Inhaltsvalidität weitgehend zufriedenstellend erreicht wird, sind die Anforderungsniveaus der Lernziele oftmals deutlich höher angesetzt als diejenigen, die in den Prüfungsaufgaben realisiert sind. Schindler (2015) stellte fest, dass Lehrende sich bei der Erstellung von Prüfungsaufgaben nicht auf vorab definierte Lehrziele beziehen und die Prüfungsaufgaben grossteils entkoppelt von den deklarierten Zielen der Veranstaltung konzipieren. Aus diesen Befunden wird deutlich, dass Dozierende in ihrer Kompetenzentwicklung begleitet und unterstützt werden müssen und die Entwicklung valider und reliabler Prüfungen nicht per se vorausgesetzt werden kann. Verständliche Anleitungen zur Konstruktion von Auswahlaufgaben oder Hilfestellungen für die Gestaltung von Beurteilungsrastern, die jeweils im Rahmen von Dozierendenversammlungen eingeführt und allen neu eintretenden Dozierenden zur Verfügung gestellt werden, bilden dafür wichtige Instrumente. Das Angebot massgeschneiderter Weiterbildungsveranstaltungen zu den identifizierten Problembereichen und individueller Beratungen durch das didaktische Zentrum unterstützen die Dozierenden bei der Ausarbeitung valider und reliabler Leistungsnachweise. Wichtig ist aber auch der Diskurs über die Gestaltung von Leistungsnachweisen. Bewährt hat sich in diesem Zusammenhang das im Rahmen der Evaluation durchgeführte Feedbackgespräch. Die evaluierten Dozierenden zeigten eine grundsätzliche Offenheit und Bereitschaft, sich mit dem von ihnen entwickelten Leistungsnachweis auseinanderzusetzen. In einem solchen Setting kann der jeweilige Leistungsnachweis im Sinne des Constructive Alignment und der Gütekriterien gemeinsam kritisch hinterfragt werden und gleichzeitig können die didaktischen Experten Ideen für eine validere Ausgestaltung des Leistungsnachweises einbringen. Anschliessend können gemeinsam Umsetzungsmöglichkeiten ausgearbeitet und diskutiert werden. In der Folge resultieren einerseits Anpassungen für den aktuellen Prüfungstermin, andererseits lassen sich Ziele für die Gestaltung künftiger Leistungsnachweise und die Formulierung von hochschulgerechten Abschlusskompetenzen formulieren. Falls grösserer Anpassungsbedarf identifiziert wurde, ist es empfehlenswert, beim nächsten Prüfungstermin ein Follow-up der Evaluation vorzusehen, wobei die didaktischen Experten idealerweise bereits bei der Entwicklung des Leistungsnachweises beratend zur Seite stehen.

Potenzial für die *Weiterentwicklung des hier beschriebenen Konzepts* besteht in verschiedenen Bereichen:

- Das Evaluationskonzept hat sich im beteiligten Departement als Bestandteil der Qualitätsentwicklung etabliert. Weil das Prüfungssystem eine wichtige Stellgrösse zur Verbesserung der Lehrqualität darstellt (Schindler, 2015), erscheint es aus der Perspektive der Gesamthochschule sinnvoll, die Evaluation summativer Lernerfolgskontrollen auf weitere Departemente zu transferieren bzw. in das Qualitätssicherungssystem der gesamten Hochschule aufzunehmen. Die breite Akzeptanz des Verfahrens bei den aktuell involvierten Personen bildet dabei eine gute Basis für eine Übertragung des Konzepts auf andere Departemente.
- Der Diskurs über die Gestaltung von Leistungsnachweisen findet derzeit nur zwischen den Didaktikexpert*innen und den Dozierenden eines Moduls statt. Ein stärkerer modulübergreifender Austausch unter den Dozierenden könnte einerseits den Diskurs über gemeinsame Standards fördern und so die Qualität des Studiengangs insgesamt erhöhen. Andererseits könnten die Dozierenden von den Ideen und Erfahrungen ihrer Kolleg*innen profitieren und so ihre Kompetenzen im Bereich des Prüfens erweitern. Allerdings erfordert ein solcher modulübergreifender Austausch erhebliche zeitliche Ressourcen und es ist unklar, inwiefern Dozierende, aber auch die Studiengangverantwortlichen hierzu bereit sind.
- Es stellt sich die Frage, wie mit der Evaluation künftig vermehrt Innovationen im Bereich des Prüfens implementiert werden können. Die digitale Transformation in der Hochschullehre bietet über den Einsatz digitaler Medien ein grosses, bislang weitgehend ungenutztes Potenzial, die Konstruktvalidität in Leistungsnachweisen zu erhöhen. So kann etwa mit dem Einsatz von bewegten Bildern (z. B. Fallvignetten) ein kompetenzorientierter Test entwickelt werden, indem die Anforderungssituation realitätsnah in die Testsituation übersetzt wird (Schindler, Bauer, Schlomske-Bodenstein & Seidel, 2015). Gerade bei Prüfenden, deren Leistungsnachweis den Anforderungen der Gütekriterien und dem Constructive Alignment entspricht, böte sich im Rahmen des Feedbackgesprächs künftig eine Diskussion an, wie der Leistungsnachweis mit Hilfe digitaler Medien und Tools innovativ weiterentwickelt werden kann. Weil die Konstruktion und Durchführung digitaler Prüfungen zusätzli-

che didaktische und infrastrukturelle Anforderungen stellen, wäre es sinnvoll, die entwickelten Ideen in Lehrprojekte zu überführen, die von Mitarbeitenden des Didaktikzentrums wissenschaftlich begleitet werden. Sofern mehrere Lehrprojekte parallel umgesetzt werden, lässt sich mit der Anwendung des Social-Support-Phasenmodells (Wahl, 2013) zusätzlich der kollegiale Austausch und der Aufbau eines Stütznetzwerks unter den Dozierenden fördern und moderieren. Eine solche Weiterentwicklung unterstützt die Dozierenden in ihrer Kompetenzentwicklung, verbessert die Qualität des Prüfungssystems und ermöglicht erste Erfahrungen mit digitalen Prüfungen. Als Ergebnis entstehen Good Practice-Beispiele sowie Handlungsempfehlungen für die Gestaltung digitaler Prüfungen. Zudem können die an Lehrprojekten beteiligten Dozierenden künftig innerhalb der Hochschule als Multiplikatoren agieren und so nachhaltig zur Hochschulentwicklung beitragen.

Leistungsnachweise steuern einerseits das Lernen von Studierenden und geben Dozierenden andererseits Feedback darüber, ob die Studierenden die angestrebten Ziele erreichen konnten. Diese Information ermöglicht es den Dozierenden das Lehr-Lernsetting entsprechend anzupassen. Das Prüfungssystem eines Studiengangs stellt damit eine bedeutsame Stellgröße zur Verbesserung der Lehrqualität und zur Studiengangentwicklung dar (Schindler, 2015). Deshalb erscheint es wichtig, Prüfungen in das Qualitätssicherungssystem an Hochschulen aufzunehmen und neben einer Lehr- auch eine Prüfungsevaluation durchzuführen. Eine solche Evaluation sollte die Kompetenzorientierung und das Constructive Alignment explizit in den Fokus nehmen und gleichzeitig als Instrument zur Kompetenzentwicklung und Innovationsförderung angelegt sein.

Literatur

- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364.
- Biggs, J. & Tang, C. (2011). *Teaching for Quality Learning at University* (4. Aufl.). Maidenhead: Open University Press.
- Frölich-Steffen, S. (2019). Einleitung. In S. Frölich-Steffen, H. den Ouden & U. Giessmann (Hrsg.), *Kompetenzorientiert prüfen und bewerten. Didaktische Grundannahmen, rechtliche Rahmenbedingungen und praktische Handlungsempfehlungen* (S. 7-10). Opladen: Barbara Budrich GmbH.

- Frölich-Steffen, S. & den Ouden, H. (2019). Hochschulprüfungen im Fokus. Vom auto-didaktischen Abschauren zu didaktisch-gereifter Prüfungskompetenz. In S. Frölich-Steffen, H. den Ouden & U. Giessmann (Hrsg.), *Kompetenzorientiert prüfen und bewerten. Didaktische Grundannahmen, rechtliche Rahmenbedingungen und praktische Handlungsempfehlungen* (S. 11-27). Opladen: Barbara Budrich GmbH.
- Hoy, W. K. & Miskel, C. G. (2013). *Educational Administration. Theory, Research, and Practice* (9th ed.). New York: McGrawHill.
- Metzger, C. & Nüesch, C. (2004). *Fair prüfen. Ein Qualitätsleitfaden für Prüfende an Hochschulen*. Hochschuldidaktische Schriften (Band 6). St. Gallen: IWP-HSG.
- Gomez, J., Kobler, D. & Nüesch, C. (2011). Die Bedeutung eines ganzheitlichen Qualitätsmanagements für die Hochschullehre. *bwp@ Berufs- und Wirtschaftspädagogik – online*, 21, 1-27. Verfügbar unter http://www.bwpat.de/ausgabe21/gomez_etal_bwpat21.pdf [01.09.2021].
- Prenzel, M. (2015). *Institutionelle Strategien zur Verbesserung der Lehre an Hochschulen: Ein Beispiel*. Festvortrag anlässlich des Jahresempfangs der Universität Bielefeld. Bielefeld: Universität.
- Schaper, N. (2021). Prüfen in der Hochschullehre. In R. Kordts-Freudinger, N. Schaper, A. Scholkmann & B. Szczyrba (Hrsg.), *Handbuch Hochschuldidaktik* (S. 87-101). Bielefeld: utb.
- Schindler, C. (2015). *Herausforderung Prüfen: Eine fallbasierte Untersuchung der Prüfungspraxis von Hochschullehrenden im Rahmen eines Qualitätsentwicklungsprogramms*. Verfügbar unter https://www.researchgate.net/publication/305800358_Herausforderung_Prufen_Eine_fallbasierte_Untersuchung_der_Prufungspraxis_von_Hochschullehrenden_im_Rahmen_eines_Qualitatsentwicklungsprogramms/references [01.09.2021].
- Schindler, C., Bauer, J., Schlomske-Bodenstein, N. & Seidel, T. (2015). Prüfungen als Indikator für Studienerfolg. In C. Berthold, B. Jorzik & V. Mayer-Guckel (Hrsg.), *Handbuch Studienerfolg* (S. 62-77). Verfügbar unter https://www.researchgate.net/publication/281117977_Prufungen_als_Indikatoren_fur_Studienerfolg [01.09.2021].
- Stefanica, F. (2013). Modulbeschreibungen – Deskriptionen realer Ansprüche oder realitätsferne Lyrik? Eine qualitative Analyse am Beispiel Höhere/Angewandte Mathematik I/II im Rahmen des Maschinenbaustudiums an ausgewählten Hochschulstandorten Baden-Württembergs. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 109(2), 286-303.
- Wahl, D. (2013). *Lernumgebungen erfolgreich gestalten* (3. Aufl.). Bad Heilbrunn: Klinkhardt.



Annette Bauer-Klebl, Dr. rer. pol., Professorin
am Zentrum für Hochschulbildung
der OST – Ostschweizer Fachhochschule.
Arbeitsschwerpunkte:
Kompetenzentwicklung und Kompetenzdiagnostik,
Digitales Lernen,
Hochschulbildung und Hochschulentwicklung

annette.bauer@ost.ch



Charlotte Nüesch, Dr. oec. HSG, Professorin
an der Pädagogischen Hochschule St. Gallen.
Arbeitsschwerpunkte:
Digitale Transformation im Bildungswesen,
Hochschulbildung,
Kompetenzdiagnostik und -entwicklung

charlotte.nueesch@phsg.ch

STICHWORT

07

Bernhard Hauser

Testing Drives Learning?

07

Bernhard Hauser

Tests beeinflussen den Alltag von Schule und Hochschule auf mannigfaltige Weise – auch in der Aus- und Weiterbildung von Lehrpersonen. In diesem Beitrag werden die Begriffe „vergleichende Tests mit Rankings“, „Teaching-to-the-Test“ und „Testing-Effekt“ erläutert und deren Wirkungen mit empirischen Befunden eingeordnet.

Vergleichstests mit Rankings, mit denen Schulen vor Ort mit anderen Regionen und Ländern verglichen werden, stoßen bei Lehrpersonen mehrheitlich eher auf Ablehnung und haben auch eher ungünstigen Einfluss auf deren Belastungsempfinden und Berufszufriedenheit (Reid, 2011; Smith & Holloway, 2020). Dies, obwohl derartige Tests als Instrumente der Qualitätssicherung und Bildungssteuerung die Leistungsfähigkeit von Schulen sicherstellen und auf diese eher positiv wirken, indem sie beispielsweise oft Grundlage sind für nachfolgende Reformen. Teaching-to-the-Test und damit das auf einen kommenden Test ausgerichtete Lernen hat Vor- und Nachteile, die stark mit der Beschaffenheit des Tests zusammenhängen. Ist der Test gut, dann hat das auf den Test ausgerichtete Lernen vor allem Vorteile. Der Einsatz von Tests während dem Lernprozess hat positive Effekte (Testing-Effekt) und kann das Lernen auf unterschiedliche Weise verbessern. Zum einen bei anspruchsvollen selektionsrelevanten und/oder qualifizierenden Tests, deren Muster bekannt sind, und die damit das vorangehende Lernen und Lehren beeinflussen, die als Learning-to-the-Test und als Teaching-to-the-Test positiv oder negativ wirken können. Insbesondere nicht-selektionsrelevante sogenannte formative Tests können das Lernen fördern, indem sie funktional als Teil einer Feedback-Kultur innerhalb transparenter Lernziele fungieren.

Vergleichende Tests mit Rankings

Im Zentrum der Kritik steht vor allem das High-stakes-Testing. Dabei handelt es sich um zentrale, internationale oder interregionale Leistungstests, die außerhalb der lokalen Schule administriert werden und deren Ergebnisse beträchtlichen Einfluss auf das unmittelbare Umfeld der überprüften Klassen haben: So fließen in der Schweiz die individuellen Ergebnisse überregionaler Vergleichstests wie Klassencockpit, einem geeichten Instrument zur Standortbestimmung der Leistungen in Mathematik und Sprache der Volksschule, in die Berechnung von Schlussnoten von Lernenden der Volksschule ein. Sie können damit

das schulische Weiterkommen der Lernenden in nachfolgende Klassen beeinflussen. In einigen Ländern haben die Ergebnisse von Vergleichstests sogar Einfluss auf Festanstellung oder Kündigung beziehungsweise Weiterbeschäftigung von Lehrpersonen. Zuweilen finden sich auch direkte Folgen für die Schulen: Diese können im nationalen oder regionalen öffentlichen Diskurs unter Druck geraten nach einer als zu wenig gut eingeschätzten Position im interregionalen oder internationalen Wettbewerb. Rangpositionen von Schulen können – je nach Land – auch einschneidende Maßnahmen nach sich ziehen in den Bereichen Finanzierung, Akkreditierung, bis hin zur Schliessung von Schulen.

Diese Vielfalt an Wirkungen findet auch ihren Niederschlag in Einstellungen und Wahrnehmungen von Lehrpersonen. So glauben angehende US-Amerikanische Lehrpersonen mit überwältigender Mehrheit, dass zu oft getestet würde und dass der Fokus auf das erfolgreiche Bestehen der Tests zu viel Einfluss auf den Unterricht habe (Reid, 2011). Es sind vor allem persönliche Erfahrungen, welche diese Skepsis gegenüber dem Testen alimentieren. Diese Studierenden des Lehramts äußern sich auch gegen ein Teaching-to-the-Test. So müsse etwa im Mathematikunterricht das konzeptuelle Verstehen im Zentrum stehen, und weniger das zu wenig verstehensorientierte Automatisieren mathematischer Regeln und Prozeduren. Letzteres werde jedoch durch Teaching-to-the-Test leider befördert. Auch bei Lehrpersonen und Lehramtsstudierenden in deutschsprachigen Ländern ist eine Skepsis diesen Tests gegenüber nicht unbekannt.

Vergleichstests dieser Art haben auch Folgen auf die Psyche von Lehrpersonen. So führen internationale Vergleichstests wie PISA zu einem zunehmenden Belastungsempfinden bei Lehrpersonen, wie Smith und Holloway (2020) in einem Überblick zu Befunden aus 33 Ländern seit dem Jahr 2013 zeigen konnten. Diese Belastungen lassen sich unter anderem darauf zurückführen, dass Lehrpersonen sich verantwortlich fühlen für das bildungsmäßige Abschneiden des eigenen Landes oder der eigenen Region. Dabei zeigte sich: je höher die Intensität der Testkultur, desto stärker waren die (negativen) Wirkungen auf die Berufszufriedenheit von Lehrpersonen. So hatte beispielsweise das Abschneiden des Landes Fürstentum Liechtenstein in der ersten PISA-Studie aus dem Jahr 2000 zu einem ganzseitigen Inserat in der wichtigsten Zeitung des Landes geführt. Darin bemängelten die einheimischen Wirtschaftsverbände das im Vergleich zu anderen eu-

ropäischen Ländern schlechte Abschneiden der Schulen des eigenen Landes mit dem Titel *Jetzt muss gehandelt werden*. Zu diesem Zeitpunkt führte ich eine schulinterne Weiterbildung in einem Dorf des Fürstentums durch und erinnere mich noch gut an die emotional und heftig geführte Debatte im dortigen Lehrpersonenzimmer. Derart öffentlich geführte Kritik löst bei Lehrpersonen viel aus, wie die Debatte um den Lehrberuf zwischen der Hochstilisierung von Lehrpersonen als Hoffnungsträger und deren Verachtung als Versager (Bastian & Combe, 2007) zeigt. Auch wenn sich Lehrpersonen nach mehr als 20 Jahren PISA inzwischen einigermaßen damit arrangiert haben, geht diese Form von öffentlichem Rechtfertigungsdruck von Bildungsinvestitionen auch heute noch nicht spurlos an ihnen vorbei. Die Folgen des Trends zu vermehrtem Wettbewerb zwischen Nationen und Regionen wurden wohl unterschätzt. Die zuweilen auf derartige Vergleichstestungen erfolgenden Bildungsreformen mögen oft zu Verbesserungen des Systems führen, beeinträchtigen aber nicht selten die Zufriedenheit von Lehrpersonen, was sich in einem Ansteigen von Stress und Burnout ausdrückt (Smith & Holloway, 2020). So führen zentrale Leistungs- und Vergleichstests bei Novizen-Lehrpersonen vermehrt zu lehrpersonen-zentrierten Lektionen (Ro, 2019), weil diese sich bei diesem Vorgehen in der Vorbereitung auf diese Tests sicherer zu fühlen. Dieses Verhalten lässt sich auch als Teaching-to-the-Test charakterisieren, eine Unterrichtsphilosophie, die oft vorschnell als dem Lernen abträglich beschrieben wird (vgl. unten).

Teaching-to-the-Test

Begegne ich früheren Kolleg*innen aus der eigenen Schulzeit, oder Kommiliton*innen aus dem Studium an der Universität, dann drehen sich unsere Gespräche nicht selten um Erinnerungen an Lehrer*innen, an Professor*innen, und um deren Ruf, der in engem Zusammenhang mit deren Anforderungen in den Prüfungen stand. Je breiter und tiefer das in deren Prüfungen eingeforderte Wissen und Können war, desto mehr wurde gelernt. Ähnliches berichten auch Studierende von heute, vor allem ehemalige. Diese – basierend auf ausgewählten subjektiven Theorien erfolgreichen – Pädagog*innen unterschieden sich aber nicht nur in der Anspruchshöhe ihrer Prüfungen, sie unterschieden sich auch darin, wie transparent sie die in den Prüfungen geforder-

ten Kompetenzen kommunizierten und wie gut sie Lernende darauf vorbereiteten. Das Lehren für die Prüfung – auch Teaching-to-the-Test genannt – ist also nichts Neues.

Teaching-to-the-Test beginnt bei Learning-to-the-Test. Vielen Lehrenden aller Stufen ist die Frage *Kommt das an der Prüfung?* sehr vertraut. Diese Orientierung Lernender und Studierender am Inhalt, den sie für eine kommende Prüfung für bedeutsam erachten, wird auch als Washback-Effekt (nach Cheng & Curtis, 2003) beschrieben. Es geht um den Einfluss, den Tests auf Lehre und Lernen haben (Alderson & Wall, 1993). „Was gemessen wird erhält Bedeutung und Wert, was wiederum die Lehre beeinflusst“ (Mc Ewan, 1995, S. 42; Übersetzung durch den Autor). Viele Lehrende, auch an Hochschulen, weisen bei sehr wichtigen Inhalten und Zusammenhängen nicht selten darauf hin, dass diese später auch Gegenstand der Prüfung sein werden. Im Kern geht es darum, wie stark die kommende – in der Regel benotete – Prüfung auf das vorangehende Lernen einen Einfluss hat. Dieser Effekt kann negativ sein, wenn der Unterricht sehr eng nur auf das Bestehen der Tests fokussiert ist, z. B. durch mechanisches Üben von im Test vorkommenden Aufgabentypen. Allerdings setzt dies einen eher schlechten Test voraus. Bei qualitativ hochwertigen Tests, wie dies bei den meisten PISA-Testaufgaben der Fall ist, kann Teaching-to-the-Test sehr ertragreich sein. Auf diesen positiven Aspekt, der vermutlich bedeutsamer ist als die häufig beklagten Nachteile, wies schon Andreas Helmke hin:

„Versteht man hingegen ‚teaching-to-the-test‘ so, dass anlässlich einer Vergleichsarbeit anspruchsvolle Aufgabentypen verstärkt thematisiert werden, und dass dabei ausreichende Gelegenheiten für horizontalen (andere Kontexte) und vertikalen (höhere Komplexität, neue Fragestellungen) Transfer gegeben werden, dann würde ich das als eine intelligente Form des Übens betrachten.“ (Helmke, 2007, 62f.).

Wichtig ist, dass relevante und anspruchsvolle Kompetenzen Teil der Prüfung sind, sonst werden diese zu wenig gelernt. Wird zum Beispiel bei Studierenden in schriftlichen Arbeiten die Kompetenz des korrekten Zitierens höher gewichtet als das Durchdringen bedeutsamer und evidenzbasierter Theorien und deren Anwendung, dann wäre das eine Folge von zu banalen Bestehenskriterien und damit ein Negativ-Beispiel für Teaching-to-the-Test. Teaching-to-the-Test ist nur dann ein Problem, wenn der Test schlecht ist.

Der Testing-Effekt

Eine besondere Variante von Teaching-to-the-Test ist der Einsatz von Übungstests. Damit können Teilbereiche einer nachfolgenden Prüfung vorher geübt werden, ohne dass das Ergebnis Konsequenzen hat für das Bestehen und Weiterkommen in der Schule. Übungstests erhöhen die Transparenz und geben Feedback zum aktuell erreichten Niveau auf dem Weg zum Können. Hierzu gibt es interessante Befunde, vor allem aus dem Bereich der Gesundheitsberufe. Auch wenn in diesem Berufsfeld Fertigkeiten und Kompetenzen oft klarer ab- und eingegrenzt werden können als Kompetenzen im Lehrberuf, so liefern Befunde dieser Studien durchaus interessante Anregungen für den Erwerb des Lehrberufs.

Der verbesserte Lernertrag durch den Einsatz von Tests zu Übungszwecken wird als Testing-Effekt bezeichnet. Danach stellen sich Lernfortschritte ein, wenn Lernende einen Übungstest zu einem Lernmaterial vor einem Schlusstest zu demselben Lernmaterial durchführen (Adesope, Trevisan & Sundararajan, 2017). Im Vergleich zu Lernenden ohne einen solchen Test schneiden erstere besser ab (vgl. auch Wood, 2009). Tests innerhalb von Lernaktivitäten – statt wie vielfach nur am Schluss des Lernprozesses – verbessern das erworbene Wissen von Lernenden stärker (Larsen, Butler & Roediger, 2008). Diese das Lernen begleitenden und nicht promotions- oder selektionsrelevanten (also unbenoteten) Tests werden auch als formative Tests bezeichnet im Gegensatz zu bilanzierenden und damit als Bestandteil von Zeugnissen fungierenden summativen Tests.

Dabei führen mehrere formative Tests zu noch besseren Ergebnissen als nur ein einzelner (Larsen et al., 2008; Roediger & Kapicke, 2006). Ebenfalls zu besseren Ergebnissen führen Testaufgaben, die verlangen, dass die Studierenden darin selbst Antworten konstruieren müssen – anstelle von Ankreuzen einer richtigen Antwort (Multiple-Choice-Test). Noch besser fallen Lernergebnisse aus, wenn auf solche Tests Feedback erfolgt. Sehr gut möglich ist dies beispielsweise mit dem digitalen Lern-Tool kahoot (genauere Informationen unter www.kahoot.com). In deren Classic-Variante können Fragen mit verschiedenen Antworten versehen werden. Den Studierenden können diese Fragen über einen Computer auf einem Screen präsentiert werden. Diese haben dann die Möglichkeit, auf ihrem Handy oder Computer

die korrekte Antwort innerhalb einer vorgegebenen Zeit als Farbfeld zu wählen. Der Wettbewerbscharakter (Punkte gibt es nur für die richtige Antwort – je schneller diese erfolgt, desto mehr Punkte gibt es) aktiviert die Konkurrenz mit den Peers und das spielerische Element. Nach jeder Frage hat die Lehrperson die Möglichkeit für ein ausführliches Feedback zu allen soeben gegebenen Antwortvarianten. Ergänzende Fragen der Studierenden vertiefen das Thema weiter. Es ist auch sehr einfach möglich, denselben Test vor und nach einer Lerneinheit einzusetzen, weil die Ergebnisse stets sofort greifbar sind und die Lehrperson keine Zeit mit Korrigieren verliert. Dieses Tool dürfte das Lernen Studierender in beachtlichem Ausmaß aktivieren.

Bei der Frage, wie nachhaltig durch wiederholte Tests erworbenes Wissen ist (Wood, 2009), interessiert vor allem, ob damit auch der Erwerb von Fertigkeiten unterstützt wird – nicht nur das Speichern von Wissen. Hierzu konnten Kromann, Jensen und Ringsted (2009; vgl. auch Pan & Rickard, 2018) zeigen, dass Lernende, die Fertigkeiten mit Hilfe von formativen Tests erworben hatten, in einem Follow-up-Test zwei Wochen später besser abschnitten, womit diese Art des Lernens auch nachhaltig ist. Besonders interessant ist der Vergleich mit dem Anfertigen von Notizen (Rummer, Schewpe, Gerst et al., 2017). Denn leider wird das wiederholte Testen zu oft mit wiederholtem Lesen verglichen. Dies, obwohl bekannt ist, dass dieses weniger wirksam ist als das Anfertigen von Notizen. Letzteres ist in sehr vielen Studienfächern sehr bedeutsam, auch in der Ausbildung von Lehrpersonen. Rummer et al. (2017) haben deshalb verglichen, ob Lernende Wissen besser behalten, wenn sie Notizen anfertigen, als wenn sie wiederholt lesen und wiederholt Tests durchführen. Unmittelbar nach der Intervention zeigte das Anfertigen von Notizen die stärksten Behaltensleistungen, eine Woche später waren die Gruppen mit angefertigten Notizen und mit wiederholtem Testen signifikant besser als die Gruppe mit wiederholtem Lesen, zwei Wochen später jedoch war nur noch die wiederholt getestete Gruppe besser als die anderen beiden. Dies ist gerade auch für angehende Lehrpersonen von Bedeutung, wird in deren Veranstaltungen doch viel Wert auf das Anfertigen von Notizen gelegt. Die Präferenz von Auszubildenden für das Anfertigen von Notizen im Vergleich zum wiederholten Testen sollte deshalb vermehrt Gegenstand von Weiterbildung werden.

Adesope et al. (2017) haben zur Frage nach der Wirkung von Übungstests eine Meta-Analyse durchgeführt und konnten zeigen, dass

Übungstests allen anderen Bedingungen überlegen waren. Die Effektgrößen fielen mehrheitlich moderat aus. Dabei war wiederholtes Testen auch wirksamer als die Wiederholung des erworbenen Inhalts.

Eine der wichtigsten Fragen für das Lehramtsstudium ist die zur Übertragbarkeit von erworbenem Wissen. Hierzu haben Pan und Rickard (2018) eine Meta-Analyse zu Transfereffekten durch Übungstests durchgeführt, die sich vor allem auf den Erwerb von Kompetenzen für Gesundheitsberufe bezog. Sie umfasste 192 Effektstärken aus 189 Studien aus mehr als 40 Jahren mit insgesamt mehr als 10.000 Teilnehmenden. Im Vergleich zu verschiedenen Kontrollbedingungen ohne Übungstests fand sich eine beachtliche Effektstärke von $d=0.4$. Besonders ertragreich waren Übungstest-Formate mit Anwendungs- und Schlussfolgerungsfragen bei Problemen, welche den Einbezug von diagnostischem Wissen erforderten. Die schwächsten Effekte fanden sich für Übungstests mit Stimulus-Response-Items, mit während dem vorangehenden Studium zwar präsentierten jedoch nicht getesteten Inhalten, und mit Problemen, die sich auf ausgearbeitete Fallbeispiele bezogen. Zudem zeigte sich eine hohe Übereinstimmung mit anfänglichen Testleistungen (Ergebnis im ersten Übungstest) sowie gute Ergebnisse für elaborierte Abruf-Expertise. Insgesamt zeigt sich, dass Übungstests das Lernen substantiell verbessern.

Schwieren, Barenberg und Dutke (2017) konnten den Testing-Effekt, wonach Übungstests zu besseren Lernerfolgen führen, ebenfalls für ertragreiches Lernen in Psychologieklassen nachweisen, beim Erwerb von Kompetenzen also, die auch für den Lehrberuf von hoher Bedeutung sind. Sie konnten diesen Testing-Effekt als robusten empirischen Befund bestätigen. Danach erleichtern Übungstests den späteren Abruf im Langzeitgedächtnis. Im Zentrum jüngerer Studien stand dabei die Übung des Abrufs aus dem Langzeitgedächtnis. Es wurden 19 Studien einbezogen, welche diesen Effekt beim Lehren und Lernen von Psychologie prüften. Das Ergebnis bestand aus total 72 Effektstärken, mit einem signifikanten und sehr beachtlichen Over-All-Effekt von der Stärke $d = 0.56$ für Übungstests in Psychologieklassen.

Bilanzierend kann festgehalten werden, dass der Einsatz von formativen Tests den Erwerb unterschiedlicher Kompetenzen nachhaltig unterstützt und gegenüber vielen Alternativen stärkere Wirkungen erzielt.

Literatur

- Alderson, J. C. & Wall, D. (1993). Does Washback Exist? *Applied Linguistics*, 14, 115-129. <http://dx.doi.org/10.1093/applin/14.2.115>
- Adesope, O. O., Trevisan, S. A. & Sundararajan, N. (2017). Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Review of Educational Research*, 87(3), 659-701.
- Bastian, J. & Combe, A. (2007). Der Lehrerberuf zwischen öffentlichen Angriffen und gesellschaftlicher Anerkennung. Alltagsbeobachtungen – professionstheoretische Erklärungen – Perspektiven der Schulentwicklung. In N. Ricken (Hrsg.), *Über die Verachtung in der Pädagogik* (S. 235-247). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Cheng, L., Curtis, A. (2004). Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning. In L. Cheng, Y. Watanabe & A. Curtis (Hrsg.), *Washback in Language Testing: Research Contexts and Methods* (pp. 3-18). Mahwah/New Jersey u. a.: Erlbaum Associates.
- Helmke, A. (2007). Guter Unterricht – nur ein Angebot? *Friedrich Jahresheft 2007*, 62-63.
- Kromann, C., Jensen, M. & Ringsted, C. (2009). The effect of testing on skills learning. *Med Educ*, 43, 21-7.
- Larsen, D. P., Butler, A. C. & Roediger, H. L. (2008). Test-enhanced learning in medical education. *Med Educ*, 42, 959-66.
- Mc Ewen, N. (1995). Educational accountability in Alberta. *Canadian Journal of Education*, 20, 27-44.
- Pan, S. & Rickard, T. C. (2018). Transfer of Test-Enhanced Learning: Meta-Analytic Review and Synthesis. *Psychological Bulletin*, 144(7), 710-756.
- Reid, P. F. (2011). Pre-service elementary teachers' mathematical beliefs and attitudes about high-stakes testing. *Dissertation. Humanities and Social Sciences, Vol 71(10 A)*.
- Ro, J. (2019). Learning to Teach in the Era of Test-Based Accountability: A Review of Research. *Professional Development in Education*, 45(1), 87-101.
- Roediger, H. L. & Karpicke, J. D. (2006). The power of testing memory: basic research and implications for educational practice. *Perspect Psychol Sci*, 132, 354-84.
- Rummer, R., Schweppe, J., Gerst, K. & Wagner, S. (2017). Is testing a more effective learning strategy than note-taking? *Journal of Experimental Psychology: Applied*, 23(3), 293-300.
- Schwieren, J., Barenberg, J. & Dutke, S. (2017). The Testing Effect in the Psychology Classroom: A Meta-Analytic Perspective. *Psychology Learning and Teaching*, 16(2), 179-196.
- Smith, W. & Holloway, J. (2020). School testing culture and teacher satisfaction. *Educational Assessment, Evaluation & Accountability*, 32(4), 461-479.
- Wood, T. (2009). Assessment not only drives learning, it may also help learning. *Medical Education*, 43, 5-6.

Bernhard Hauser, Prof. Dr. phil.
an der Pädagogischen Hochschule St. Gallen.
Arbeitsschwerpunkte:
Lehren und Lernen,
Lernwirksamkeit von Spiel,
Bildungsforschung bei 3- bis 10-Jährigen



bernhard.hauser@phsg.ch

REZENSION

Reh, S., Bühler, P., Hofmann, M. & Moser, V. (2021). *Schülersauslese, schulische Bewertung und Schülertests 1880-1980*. Verlag Julius Klinkhardt. 277 Seiten. ISBN 978-3-7815-2458-3

Verschiedene Formen der Begutachtung, der Testung und der Leistungsmessung sind wesentliche Aufgaben, die die Institution Schule erfüllen muss. Auch wenn die Praktiken, die dort zum Tragen kommen, fortlaufender Kritik ausgesetzt sind, sind sie doch konstitutiv für die uns bekannten Formen schulischer Bildung.

Diskussionen und kontroverse Debatten um Formen des *Prüfens* und *Bewertens* sind nicht nur heute – insbesondere vor dem Hintergrund der Bemühungen um Inklusion – hochaktuell, sondern lassen sich schon in den historischen Diskursen der letzten Jahrhunderte finden. Durch die wachsenden Versuche der Beschulung von Schüler*innen, die von der erwarteten Leistungsnorm abweichen, intensivierten sich diese Debatten. Sie umfassen, heute wie damals, nicht nur Praktiken der fachlichen Leistungsbewertung, sondern auch Praktiken der Diagnostik und damit verbundene Prozesse von In- und Exklusion.

Der Band *Schülersauslese, schulische Bewertung und Schülertests 1880-1980*, herausgegeben von Sabine Reh, Patrick Bühler, Michèle Hofmann und Vera Moser, widmet sich aus bildungshistorischer Perspektive den unterschiedlichsten Aspekten pädagogischer und schulpolitischer Debatten um Praktiken des Beobachtens, Prüfens und Beurteilens von Schüler*innen und den mit diesen verbundenen Verfahren und ihren Folgen.

Mit seinen insgesamt 15 Beiträgen gibt der Band Einblicke in unterschiedlichste Dimensionen der Diskussionen des vergangenen Jahrhunderts und beleuchtet, wie die Herausgeber*innen in der Einleitung formulieren, das „[...] entstehende Zusammenspiel zwischen der Etablierung und Differenzierung institutionalisierter Bildungsangebote, den Praktiken der Beobachtung, Prüfung, Testung und Begutachtung sowie den Diskursen über Begabung, Intelligenz und Leistung.“ (S. 9) So widmen sich mehrere Beiträge den Praktiken und Begründungsmustern des Ausschlusses von Kindern, die als „schwachbegabt“ (z. B. S. 65), „bildungsunfähig“, „anormal“, „idiotisch“ (z. B. S. 67) oder „schwachsinnig“ (z. B. S. 241) bezeichnet wurden, und den damit verbundenen Bemühungen um den Ausbau eines Hilfsschulsystems und den sich immer weiter ausdifferenzierenden Praktiken der Diagnostik.

Patrick Bühler greift diese Thematik auf mit dem Beitrag *Komplett pessimistisch eingestellt Hilfe und Heilung in der Schweizer Sonderpädagogik zu Beginn des 20. Jahrhunderts*, Michèle Hofmann mit seinem Beitrag *Grenzziehungen – Praktiken der Kategorisierung geistig anormaler Kinder um 1900 in der Schweiz*, Nadja Wenger mit ihrem Beitrag *Ihr gebt mich fort, weil ihr mich nicht gern habt. Gutachten der St. Galler Fürsorgestelle für Annormale in den 1940er Jahren* und auch Jona Garz, Vera Moser und Stefan Wünsch unter dem Titel *Die Kielhorn-Rede: Ursprungsmythos der deutschen Hilfsschulen*, sowie Michaela Vogt im letzten Beitrag des Bandes über *Das Hilfsschulnahmeverfahren als Grenzzone der Schülerschulung in BRD und DDR*.

Bühler und Hofmann nehmen zu Anfang ihrer Beiträge jeweils Bezug auf den Beginn des sukzessiven Ausbaus eines Hilfsschulsystems in der Schweiz. Bühler betont hierbei insbesondere die Institutionalisierung der Heilpädagogik und geht dann weiter auf die schon damals existierende Problematik des eher medizinisch verorteten Begriffs des „Heilens“ ein (vgl. S. 87). Hofmann konzentriert sich in seinem Beitrag auf die Ausdifferenzierung verschiedener Beschulungs- oder auch Verwahrungskonzepte für von der Norm abweichende Kinder und Jugendliche und erläutert dabei dezidiert die bis heute existierende Problematik der Herstellung einer Norm, von welcher dann eine Abweichung konstruiert wird.

Garz, Moser und Wünsch bereichern die Debatte um die Gründung von Hilfsschulen mit einem Beitrag, der die bisherige Historiographie des Sonderschulwesens kritisch beleuchtet. Die bekannten Erklärungsmuster der Gründung von Hilfsschulen, insbesondere die wachsende Industrialisierung und eine fortschrittsoptimistische Form der Geschichtsschreibung, werden durch die Autor*innen erweitert um „[...] ein vielschichtiges Netzwerk von zeitgenössischen Wissensordnungen und protestantischen Verheißungen [...]“ (S. 42). Sie kommen zu dem Schluss, dass pädagogische Erwägungen kaum eine Rolle spielten und psychiatrische und christliche Deutungsmuster die relevanten Lieferanten der neuen Wissensbestände waren, die zur Konstruktion von Hilfsschulen und somit auch dem Hilfsschulkind beitrugen (vgl. S. 42). Zentral sind in diesen Beiträgen immer wieder die Diskussionen um Verfahren der Feststellung von Abweichung sowie der anschließende Umgang der besonderen Beschulung – Fragen, die auch im heutigen Schulsystem trotz aller inklusiven Bestrebungen als nicht endgültig beantwortet betrachtet werden können.

Eine besondere Praktik der *Testung* beschreibt Jona Garz in seinem Beitrag *Schriftproben von schwachsinnigen resp. Idiotischen Kindern. Testwissen zwischen Psychatrie und Pädagogik um 1900*. Garz illustriert in diesem Beitrag, wie Schreiben als ein Diagnoseinstrument genutzt wurde, wobei insbesondere die Anordnung der Buchstaben und weniger der Inhalt zählte (vgl. S. 52). Die Handschrift wurde dabei als „materialisierter Ausdruck der Störung des Gehirns“ (S. 52) betrachtet. Mit Hilfe der Analyse von Schriftproben wurde davon ausgegangen so zwischen „verschiedenen Arten des Schwachsinnns zu differenzieren und den Schweregrad der Idiotie“ messen zu können. Diese Form des Tests, der als ein objektives Diagnosekriterium (vgl. S. 58) galt, zeigt exemplarisch, welche hohen Stellenwert standardisierte Verfahren hatten und der Beitrag verweist auch explizit darauf, welchen Erfolg diese auch in Teilen bis heute im Schulsystem haben.

Anschlussfähig an die immer wieder aufkeimenden Debatten um die Möglichkeit einer neutralen oder objektiven Bewertung, insbesondere von Aufsätzen, zeigt sich Kerrin von Engelhardts Beitrag *Der papierenen Drache – Der Reifeprüfungsaufsatz zwischen 1890 und 1930*. Von Engelhardt beschreibt hier die in der Praktik der Reifeprüfungsaufsätze zu beobachtende „hybride Struktur“ (S. 172) der Bewertung, bei der sowohl fachspezifische Kriterien herangezogen wurden als auch übergeordnete, wie beispielsweise ein Ermessensspielraum, der sich aus weiterführendem Wissen der Fachlehrer*in über die zu bewertenden Schüler*innen ergab. Die Form des Changierens zwischen übergeordneten und individuellen Bewertungsmustern und die von Von Engelhardt aufgegriffene Thematik des Problems der individuellen Präferenzen und Sympathien, wird somit als schulische Herausforderung erkennbar, mit der sich pädagogisches Fachpersonal nicht erst in den letzten Jahrzehnten konfrontiert sah.

Einen bereichernden Blick über die Grenzen des deutschsprachigen Raums hinaus gibt Fanny Isensee in ihrem Beitrag *Intelligence tests were given in order to obtain a basis for classifying the pupils – Die reclassification Projects in New York City in den 1920er-Jahren*. Anhand des Beispiels Brooklyn diskutiert Isensee die Nutzung von Intelligenztests zur Einteilung von Schüler*innen in bestimmte Klassenstufen. Besonders aufschlussreich und neu ist Isensees Teilfragestellung, inwiefern die Ergebnisse aus solchen Tests auch Rückwirkung auf die Institution Schule und ihre Ausgestaltung hatten. Isensee kommt zu dem Ergebnis, dass Resultate von Intelligenztests auch genutzt wur-

den um die Schulen effizienter und wirtschaftlicher (vgl. S. 239) zu gestalten. Isensees Analysen zeigen zudem, was kaum verwundern mag, aber doch wieder ein Augenmerk auf die Beharrlichkeit solcher und ähnlich gelagerter Annahmen richtet, dass Schüler*innen mit Migrationshintergrund immer mit dem Vorwurf konfrontiert waren, schlechtere Schulleistungen zu zeigen und somit auch schlechter in IQ-Tests abzuschneiden (vgl. S. 238).

Neben diesen diversen Beiträgen, die sich eher der Problematik des Umgangs mit *negativ* von der Leistungsnorm abweichenden Schüler*innen befassen, greift Susanne Schregel in ihrem Beitrag *Übernormalen Pädagogik und Begabenschulen zwischen Kaiserreich und Weimarer Republik* die Thematik um Hochbegabte auf. Zentrales Thema des Beitrages sind Diskussionen um die Überführung von Hochbegabten in „Klassen für Übernormale“, „Sonderschulen für hervorragend Befähigte“ oder „Genie-Schulen“ (S. 135). Hierbei stellt sie durch die Analyse von zeitgenössischen pädagogischen und psychologischen Fachpublikationen heraus, dass die Verbreitung der Intelligenzdiagnostischen Verfahren vermehrt dazu führte, dass sowohl positive als auch negative Abweichung als fast gleichartige „Rand“-Phänomene betrachtet wurden (vgl. S. 147). Somit zeigt sie, anknüpfend an auch heute viel rezipierte kritische Positionen, dass diagnostische Verfahren die Gefahr bergen, eine künstliche Leistungsnormalität zu konstruieren, die jegliche Form der Abweichung als unnormal und als potenziell störend kennzeichnet.

Dass sich psychologische Tests und Diskussionen um Normalität und Abweichung nicht nur darauf beschränkten *innerhalb* des Schulsystems homogene Leistungsgruppen herzustellen oder Schüler*innen eine als angemessen angesehene Förderung zukommen zu lassen, zeigt der Beitrag *Das Kind vor verfehlter Wahl geschützt. Die Einführung eines berufspsychologischen Schülerbeobachtungsbogens in Berliner Schulen, 1917-1923* von Johanna Lerch. Sie erläutert, wie in der Weimarer Republik die Verbesserung der Berufswahl jugendlicher Schulabgänger*innen zu einem politischen Thema wurde und große Erwartung in psychotechnische diagnostische Verfahren gesetzt wurde, um potenzielle Arbeitskräfte ökonomisch nutzbringend auf verschiedenen Wirtschaftszweige zu verteilen (vgl. S. 209). Schon damals standen einige Pädagog*innen und Psycholog*innen solchen Tests kritisch gegenüber und stufen diese als „ungenügend, oberflächlich und zufällig“ ein (S. 211), weshalb sie alternative Formen der Beob-

achtung vorschlugen. Anhand des Vergleiches verschiedener Beobachtungsbögen zeigt Johanna Lerch auf, „[...] dass die Abwendung von der experimentellen Psychologie hin zur Typenlehre und später zum psychologischen Test [...] sich in stetiger Auseinandersetzung mit ihren Anwender*innen weiterentwickelte.“ (S. 222) Diese Ergebnisse werfen die Frage auf, inwiefern auch heute Lehrkräfte als Anwender*innen von Tests die Ausgestaltung dieser mitbestimmen und somit als relevante Akteur*innen im Diskurs um Diagnoseverfahren angesehen werden müssen.

Die hier exemplarisch kurz vorgestellten Beiträge zeigen schon die Vielfältigkeit der Möglichkeiten, sich aus historischer Perspektive dem Phänomen der Schülersauslese und damit verbundenen Beurteilungs- und Testpraktiken zu nähern auf und die diskursiven Parallelen zu heutigen Debatten scheinen eklatant.

Obwohl einige Beiträge des Bandes weniger bekannte Praktiken von Leistungsbewertung und Tests aufgreifen, setzt er das Thema der Intelligenztests und die mit diesen verknüpften diskursiven Verstrickungen sehr zentral, sodass sich dieses im Vergleich zu anderen Praktiken der Diagnose und Leistungsmessung etwas überrepräsentiert zeigen. Weitere tiefere Einblicke in die Umsetzung der spezifisch methodisch unterschiedlichen Verfahren, wie sie beispielsweise der Beitrag von Garz zu Schriftproben oder von Von Engelhardt zum Reifeprüfungsaufsatz geben, wären ergänzend sehr lohnenswert gewesen.

Nichtsdestotrotz zeichnen sich die im Band versammelten Beiträge, sowohl hinsichtlich der zeitlichen Zuschnitte als der gewählten thematischen Einblicke durch eine große Heterogenität aus – ohne dabei einen roten Faden vermissen zu lassen. Der Band schafft es darüber hinaus, insbesondere für Leser*innen, die über aktuelle Debatten um das Schulsystem informiert sind, anhand von unterschiedlichen Aspekten die lange zurückreichende Kontinuität um ein Streben nach einer Homogenisierung von Lerngruppen (vgl. S. 101) zu illustrieren. Das Zusammenspiel von Inklusion als gesellschaftlicher Aufgabe und die Bearbeitung dieser mit schulisch oder weiter gefassten pädagogischen Mitteln, ist dabei immer mit zentral gesetzt.

In besonderem Maße zeigt sich aber auch immer wieder, dass konträre Positionen und Aushandlungsprozesse rund um Praktiken der Messung, Diagnostik, Prüfung und Bewertung, welche eng mit Ein- und Ausschlussprozessen verbunden waren und sind, existierten. Zudem

werden auch Ambiguitäten und Paradoxien rund um exkludierende, integrierende oder inkludierende schulische und pädagogische Prozesse aufgezeigt, in denen sich immer wieder fragmentarische Parallelen zu heutigen Debatten um das Inklusionsparadoxon oder auch das Etikettierungs- und Ressourcendilemma wiederfinden, was den Band auch für eine nicht genuin bildungshistorisch orientierte Leser*innenschaft äußerst interessant macht.

Friederike Thole, wiss. Mitarbeiterin
am Institut für Erziehungswissenschaft –
Fachgebiet Historische Bildungsforschung
der Ruhr-Universität Bochum

friederike.thole@rub.de

AGENDA

(Stand 01.03.2022)

5. Tagung Fachdidaktiken

Informationen: <https://www.swissuniversities.ch/themen/fachdidaktiken/tagung-fachdidaktiken>

08.-09.04.2022,
Fachhochschule Südschweiz
Locarno, Schweiz

XVII Internationale Tagung der Deutschlehrerinnen und Deutschlehrer

Informationen: <https://www.idt-2021.at/>

15.-20.08.2022
Universität Wien,
Österreich

ADLD Conferences 2022

Informationen: <https://sgl-online.ch/events/scientific-conference-advancing-digital-learning-discourse-in-teaching-teacher-education-and-teachers-professional-development/>

04.-09.09.2022,
Universität Zürich,
Schweiz

VASOM: Vienna Autumn School of Methods

Informationen:
<https://vasom.univie.ac.at/>

08.-10.09.2022,
Universität Wien,
Österreich

SGBF-Kongress 2022

Informationen: <https://www.sgbf.ch/sgbf/aktuelles/news/details/news/sgbf-kongress-2022>

12.-13.09.2022,
Universität Lausanne,
Schweiz

20.-23.09.2022,
PPH Augustinum und
PH Steiermark
Österreich

ÖFEB Kongress 2022

Informationen:

<https://www.oefeb.at/veranstaltungen/view/99>

24.09.2022,
PH Zürich,
Schweiz

2. Internationale Tagung

„Achtsamkeit in Schule und Bildung“ 2022

Informationen: <https://phzh.ch/achtsamkeit2022>

26.11.2022,
PH Zürich,
Schweiz

Tagung SpielRäume 2022

Informationen: [https://phzh.ch/de/Forschung/](https://phzh.ch/de/Forschung/Forschungsveranstaltungen/veranstaltung/Tagung-SpielRaeume-2022-v144500878.html)

[Forschungsveranstaltungen/veranstaltung/](https://phzh.ch/de/Forschung/Forschungsveranstaltungen/veranstaltung/Tagung-SpielRaeume-2022-v144500878.html)

[Tagung-SpielRaeume-2022-v144500878.html](https://phzh.ch/de/Forschung/Forschungsveranstaltungen/veranstaltung/Tagung-SpielRaeume-2022-v144500878.html)

CALL FOR ABSTRACTS

Im journal für lehrerInnenbildung werden vier Themenhefte pro Jahr herausgegeben. Ab 2017 werden mehrere Beiträge jedes Heftes nach einem offenen Call ausgewählt. Weitere Beiträge werden wie bis anhin von den für das jeweilige Heft verantwortlichen Herausgeber*innen eingeworben. Die abschließende Auswahl liegt bei den für das Themenheft verantwortlichen Herausgeber*innen. Kriterien für offenen Call: Die Beiträge sollen die Aus- und Weiterbildung von Lehrpersonen fokussieren bzw. einen Zusammenhang zur Lehrer*innenbildung aufweisen und zum Heftthema passen.

Das jlb publiziert nur Erstveröffentlichungen. Es sind verschiedene Beitragstypen möglich: Überblicksartikel, Forschungsberichte, Erfahrungsberichte und Konzeptbeschreibungen. Die Textlängen bewegen sich je nach Beitragstyp zwischen 10.000 und 20.000 Zeichen (mit Leerzeichen).

Im offenen Call wird ein Abstract (max. eine A4-Seite) zum Thema des Heftes erwartet. Darin enthalten sind Bezüge zur wichtigsten Forschungsliteratur und eine Angabe zum Beitragstyp.

Per mail an den/die jeweilige Herausgeber*in des Hefts.

02/2023 Service Learning in der Lehrer*innenbildung

(Herausgeber*innen: Jennifer Paetsch, jennifer.paetsch@uni-bamberg.de und Katharina Resch, katharina.resch@univie.ac.at)

Service Learning gilt als ein hochschuldidaktischer Lehransatz, der Studium, Lernen, Engagement, Berufsvorbereitung und gesellschaftliche Bedarfe verzahnt und so Theorie und Praxis verbindet. Im Service Learning werden gesellschaftliches Engagement von Studierenden (*service*) mit paralleler Reflexion von Erfahrungen (*learning*) zusammengeführt. Service Learning stellt das Engagement von Studierenden außerhalb der Hochschule für das Gemeinwohl in den Vordergrund, was zur Öffnung der Hochschulen beitragen soll.

Service Learning ermöglicht es Studierenden des Lehramts, an einer organisierten Aktivität – etwa an Schulen, in der Gemeinde oder in sozialen Organisationen – mitzuwirken und diese Erfahrung in einer dazugehörigen Lehrveranstaltung zu reflektieren. Durch die Mitwirkung von Studierenden in einer organisierten Dienstleistung in Schulen oder anderen pädagogischen Handlungsfeldern hat der Ansatz auch eine Funktion in der Berufsorientierung inne. Studierende lernen so das spätere Berufsfeld kennen.

Der Begriff des Service Learning als Lehransatz mag vielen Lehrenden in der Lehrer*innenbildung bisher unbekannt sein, während andere bereits seit vielen Jahren damit arbeiten. Das Themenheft soll dazu beitragen, Wissen, Erfahrungen und Konzepte zu diesem Lehransatz im deutschsprachigen Raum voranzubringen. Beiträge des geplanten Themenheftes können den Mehrwert von Service Learning in der Lehrer*innenbildung untersuchen, kritische Aspekte zum Einsatz von Service Learning beleuchten oder Erfahrungen aus der praxisorientierten Lehre sowie Typen und Umsetzungsformen von Service Learning in der Lehrer*innenbildung darstellen.

Deadline für den offenen Call: Juni 2022

Erscheinungstermin: Juni 2023

03/2023 Bildung für Nachhaltige Entwicklung

(Hrsg.: Caroline Bühler, caroline.buehler@phbern.ch)

Bildung für nachhaltige Entwicklung (BNE) wird angesichts der aktuellen ökologischen und sozialen Herausforderungen als eine dringliche Aufgabe an die Lehrer*innenbildung herangetragen. Die Hochschulen reagieren mit einer Vielfalt von Konzepten und Umsetzungsaktivitäten und es wird viel unternommen, um BNE auch in die Schulen zu tragen. Für das Heft interessant ist zunächst die Frage nach dem Bildungsverständnis, das in die Lehrer*innenbildung im Zusammenhang mit BNE-Konzepten transportiert wird. In welchen BNE-Diskurs ordnen sich die Studiengänge ein, welche Zugänge wählen sie? Oder auch: Wie verankern sie BNE in den Curricula?

Neben theoretisch-konzeptionellen Beiträgen möchten wir auch Forschungsbefunde zu Formen der Umsetzung von BNE dokumentieren – sowohl in Bezug auf die Unterrichtsfächer als auch auf die Erziehungs- und Sozialwissenschaften.

Schliesslich sind wir neugierig auf konkrete Umsetzungserfahrungen in der Praxis der Lehrer*innenbildung: Welche inhaltlichen Ausrichtungen und welche Arbeitsweisen werden gewählt? Wie werden Studierende in die Konzeption der Themen eingebunden?

Deadline für den offenen Call: September 2022

Erscheinungstermin: September 2023

Wir freuen uns auf zahlreiche Eingänge!
Für die jlb-Redaktion: Stefanie Thurnheer

Testen und Bewerten

journal für lehrerInnenbildung

no. 1/2022

EDITORIAL

BEITRÄGE

01

Mündliche Prüfungen im Kontext des Forschenden Lernens.
Diskurs und Praxis

02

Analysekompetenz von Studierenden mit einem Video-Portfolio beurteilen

03

Die Beurteilung angehender Lehrer*innen im Referendariat.
Relevanz, Kontext und erste Explorationen

04

Einblicke in die soziale Praxis des Prüfens an Hochschulen.
Innen(ein)sichten

05

Mythos Eignungsprüfung.
Kunst – Leistung – hochschulische Prüfungspraxis

06

Prüfungsevaluationen und damit verbundene Feedbackgespräche:
Wie sie zur Hochschulentwicklung beitragen können

STICHWORT

07

Testing Drives Learning?

REZENSION

AGENDA

CALL FOR ABSTRACTS