pedocs
Open Access Erziehungswissenschaften

DIPF

Kärner, Tobias; Bonnes, Caroline; Schölzel, Christian

Bewertungstransparenz im Referendariat

Zeitschrift für Pädagogik 65 (2019) 3, S. 378-400

Quellenangabe/ Reference:
Kärner, Tobias; Bonnes, Caroline; Schölzel, Christian: Bewertungstransparenz im Referendariat - In: Zeitschrift für Pädagogik 65 (2019) 3, S. 378-400 - URN: urn:nbn:de:0111-pedocs-239481 - DOI: 10.25656/01:23948

https://nbn-resolving.org/urn:nbn:de:0111-pedocs-239481
https://doi.org/10.25656/01:23948

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

http://www.juventa.de

Nutzungsbedingungen

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

Terms of use

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

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

Kontakt / Contact:

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

Mitglied der Leibniz-Gemeinschaft

D 7484

ZEITSCHRIFT FÜR PÄDAGOGIK

Heft 3 Mai/Juni 2019

■ *Aktuelle Forschungsdiskurse*

… zur inklusiven Schule

… zur LehrerInnenbildung und zum Lehrberuf

… zur Digitalisierung von Bildung

BELTZ JUVENTA

ZEITSCHRIFT FÜR PÄDAGOGIK
Jahrgang 65 – Heft 3
Mai/Juni 2019

Inhaltsverzeichnis

Aktuelle Forschungsdiskurse

… zur inklusiven Schule

Adina Küchler/Alina Ivanova
Anerkennung als Voraussetzung für inklusives und migrationspädagogisches Handeln in der Schule 315

Jan Jochmaring
Übergänge von Schüler/innen mit sonderpädagogischem Förderbedarf in die Berufsausbildung. Eine Auswertung von Sekundärstatistiken 335

… zur LehrerInnenbildung und zum Lehrberuf

Stefan Klemenz/Johannes König
Modellierung von Kompetenzniveaus im pädagogischen Wissen bei angehenden Lehrkräften. Zur kriterialen Beschreibung von Lernergebnissen der fächerübergreifenden Lehramtsausbildung 355

Tobias Kärner/Caroline Bonnes/Christian Schölzel
Bewertungstransparenz im Referendariat 378

Colin Cramer/Martin Harant/Samuel Merk/Martin Drahmann/ Marcus Emmerich
Meta-Reflexivität und Professionalität im Lehrerinnen- und Lehrerberuf 401

… zur Digitalisierung von Bildung

Sigrid Hartong
Bildung 4.0? Kritische Überlegungen zur Digitalisierung von Bildung als erziehungswissenschaftliches Forschungsfeld 424

Diskussion

Fritz Bohnsack
Noch einmal „Anders als gedacht". Ein Kommentar zu Michael Knolls Dewey-Aufsatz in der Zeitschrift für Pädagogik (2018, Heft 5) 445

Besprechungen

Johannes Drerup
Anke Redecker: Das kritische Selbst. Bildungstheoretische Reflexionen im Anschluss an Hugo Gaudig, Marian Heitger, Käte Meyer-Drawe und Immanuel Kant 451

Heinz-Elmar Tenorth
Hartmut Rosa/Wolfgang Endres: Resonanzpädagogik. Wenn es im Klassenzimmer knistert (2. Aufl., mit einem Nachwort von Reinhart Kahl).
Hartmut Rosa/Wolfgang Endres/Jens Beljan: Resonanz im Klassenzimmer. 48 Impulskarten zur Resonanzpädagogik mit 16-seitigem Booklet.
Hartmut Rosa/Claus G. Buhren/Wolfgang Endres: Resonanzpädagogik und Schulleitung. Neue Impulse für die Schulentwicklung.
Hartmut Rosa: Resonanz. Eine Soziologie der Weltbeziehung 453

Berno Hoffmann
Eckart Conze/Susanne Rappe-Weber (Hrsg.): Die deutsche Jugendbewegung. Historisierung und Selbsthistorisierung nach 1945 (Jugendbewegungen und Jugendkulturen, Bd. 14) 458

Ulrich Heimlich
Tanja Sturm/Monika Wagner-Willi (Hrsg.): Handbuch schulische Inklusion 461

Jane Weiß
Albert Scherr/Lena Sachs: Bildungsbiographien von Sinti und Roma. Erfolgreiche Bildungsverläufe unter schwierigen Bedingungen 463

Dokumentation

Pädagogische Neuerscheinungen 467

Impressum U3

Table of Contents

Current Research Discourses

… on Inclusive Schools

Adina Küchler/Alina Ivanova
Recognition as a Prerequisite for Inclusive and Migration Pedagogical Action in School 315

Jan Jochmaring
Transitions of Students with Special Educational Needs into Vocational Training: an evaluation of secondary statistics 335

… on Teacher Training and the Teaching Profession

Stefan Klemenz/Johannes König
Modeling Proficiency Levels of Pre-Service Teachers' Pedagogical Knowledge: On the criterial description of cross-subject teacher education outcomes 355

Tobias Kärner/Caroline Bonnes/Christian Schölzel
Assessment Transparency in Teacher Training 378

Colin Cramer/Martin Harant/Samuel Merk/Martin Drahmann/ Marcus Emmerich
Meta-Reflexivity and Professionalism in the Teaching Profession 401

… on Digitalisation of Education

Sigrid Hartong
Education 4.0? Critical reflections on the digitalisation of education as an educational research field 424

Discussion

Fritz Bohnsack
Once again "Different than thought". A comment on Michael Knoll's article on Dewey in this periodical (2018, 5) 445

Book Reviews 451

New Books 467

Impressum U3

Tobias Kärner/Caroline Bonnes/Christian Schölzel

Bewertungstransparenz im Referendariat

Zusammenfassung: Bestehende Studien verweisen auf Optimierungspotenziale hinsichtlich der Bewertungspraxis in der zweiten Phase der Lehrer/-innenbildung (vgl. Strietholt & Terhart, 2009). Insbesondere ist sehr wenig darüber bekannt, wie transparent die Bewertung im Referendariat seitens angehender Lehrpersonen wahrgenommen wird. Vor diesem Hintergrund und in Ermangelung eines entsprechenden Instruments stellen wir in unserem Beitrag einen Fragebogen zur Erfassung der wahrgenommenen Bewertungstransparenz vor. In einem ersten Schritt wurden (1) nachvollziehbare und offen kommunizierte Leistungserwartungen und Bewertungskriterien sowie (2) konstruktives Feedback als zentrale Dimensionen einer transparenten Leistungsbewertung literaturbasiert herausgearbeitet. Die beiden Dimensionen in Kombination mit insgesamt fünf Bewertungsanlässen (z. B. Prüfungslehrproben, mündliche Prüfungen) dienten im zweiten Schritt als Grundlage für die Formulierung entsprechender Items. Die faktoranalytische Prüfung stützt die Annahme der beiden Dimensionen, welche je Bewertungsanlass mit jeweils vier Items erfasst werden. Die gefundenen Skalenwerte bewegen sich deskriptiv im mittleren Bereich, was weder auf eine besonders schlechte, noch auf eine besonders gute Ausprägung der wahrgenommenen Bewertungstransparenz in der zugrundeliegenden Stichprobe hinweist. Auffallend sind jedoch die durchgängig hohen Standardabweichungen. Implikationen ergeben sich für den von uns entwickelten Fragenbogen dergestalt, dass dieser u. a. im Rahmen des Vorbereitungsdienstes zu Evaluationszwecken eingesetzt werden kann, um auf dieser Grundlage Faktoren zu identifizieren, welche in spezifischen Fällen zu einer besonders guten bzw. schlechten Bewertungstransparenz beitragen, um im letzteren Fall wiederum Ansatzpunkte für eine Optimierung der zweiten Phase der Lehrer/-innenbildung zu identifizieren.

Schlagworte: Leistungsbewertung, Referendariat, Bewertungskriterien, Feedback, Transparenz

1. Ausgangslage und Zielstellung

Der Vorbereitungsdienst für das Lehramt an beruflichen wie an allgemeinbildenden Schulen wird mitunter als psychisch herausfordernd und anstrengend erlebt, u. a. deshalb, da die Referendare/-innen in dieser Zeit einer Vielzahl von Bewertungsanlässen ausgesetzt sind (vgl. Strietholt & Terhart, 2009). Bundesland- und schulartübergreifend finden sich diesbezüglich insbesondere Seminarveranstaltungen, Unterrichtsbesuche bzw. Lehrversuche, Prüfungslehrproben, ein Kolloquium und mündliche Prüfungen sowie die Seminararbeit, wobei die Lehramtsanwärter/-innen ihre Fähigkeiten und Kenntnisse immer wieder aufs Neue unter Beweis stellen müssen und auf dieser Grundlage entsprechend bewertet werden (vgl. KMK, 2012; KMK, 2013).

Eine transparente Leistungsbewertung setzt hierbei zum einen durchschaubare, nachvollziehbare und verständlich formulierte sowie vor einer Prüfungssituation offen kom-

munizierte Leistungserwartungen und Bewertungskriterien voraus, wobei verbindliche inhaltliche Ziele zwischen Bewertenden und Bewerteten festzulegen und einzuhalten sind. Zum anderen setzt sie einen nachvollziehbaren und verständlichen Rückmelde- bzw. Feedbackprozess sowie einen konstruktiven Umgang mit Fehlern voraus, in dessen Rahmen es möglich ist, einem Prüfling ihre/seine Stärken und Schwächen aufzuzeigen. *Transparente Leistungserwartungen und Bewertungskriterien* einerseits sowie *konstruktives Feedback* andererseits werden demnach als *zentrale Dimensionen* einer transparenten Leistungsbewertung angesehen (vgl. Hattie & Timperley, 2007; Klein, 2009; Meyer, 2004; Oser, Hascher & Spychiger, 1999; Paradies, Wester & Greving, 2005; Winne & Butler, 1994). Demgegenüber können intransparente Bedingungen und die Unkenntnis darüber, was in der Leistungsbewertungssituation von einem erwartet wird, mit dem Erleben von Stress und Prüfungsangst einhergehen (vgl. Kaluza, 2015; Rost & Schermer, 2010). Eine möglichst stress- und angstfreie Atmosphäre kann unter anderem dadurch geschaffen werden, indem der genaue Ablauf einer Prüfung bekannt gegeben wird, die Prüfungsinhalte transparent kommuniziert und angemessen dargestellt sowie die erwarteten Ansprüche und Beurteilungsmaßstäbe bekannt gegeben werden (vgl. Beiner, 1982). Beispielsweise zeigen Drüge, Schleider und Rosati (2014) in ihrer Studie zu psychosozialen Belastungen im Referendariat, dass die wahrgenommene Vorhersehbarkeit im Arbeitsprozess belastungsreduzierend wirken kann. Da Lehramtsanwärter/-innen (vgl. Klusmann et al., 2012) wie auch fertig ausgebildete Lehrpersonen (vgl. Rudow, 2011; Schaarschmidt, 2011) über besondere berufsbezogene Belastungen berichten und vor allem in Bezug auf psychische und psychosomatische Erkrankungen als gesundheitlich gefährdete Personengruppen anzusehen sind (vgl. Scheuch, Haufe & Seibt, 2015), erscheint eine genauere Betrachtung belastungsreduzierender sowie zufriedenheitssteigernder Arbeits- bzw. Ausbildungsmerkmale hierbei als besonders angebracht.

Das Thema „transparente Leistungsbewertung" wird bereits seit Jahrzehnten im Zusammenhang mit der Bewertung von Schüler/-innenleistungen beschrieben, diskutiert und auch in der Lehramtsausbildung behandelt (vgl. Ingenkamp, 1974; Jürgens, 1998; Klein, 2009; KMK, 2014; Rheinberg, 2001), wohingegen Studien, welche die Leistungsbeurteilung von Lehramtsanwärter/-innen im Referendariat betrachten, deutlich die Ausnahme darstellen (vgl. Strietholt & Terhart, 2009). Bestehende Studien lassen hierbei Zweifel an einer durchgängig transparenten Bewertungspraxis aufkommen. So berichten beispielsweise Döbrich und Abs (2008), dass 61.6 % der befragten Lehramtsanwärter/-innen der Aussage, dass die Bewertung von Referendaren/-innen einheitlichen Kriterien folge, *eher nicht oder gar nicht* zustimmen. Demgegenüber stimmen jedoch 70.3 % der mit der Ausbildung von Referendaren/-innen betrauten Lehrpersonen dieser Aussage *eher oder voll* zu. Strietholt und Terhart (2009) veranschaulichen in ihrer Studie eindringlich vorhandene Defizite im Zusammenhang mit den Bewertungsinstrumenten und -prozessen in der zweiten Phase der Lehrer/-innenbildung. So bezieht sich der Großteil der gesichteten Beurteilungsinstrumente lediglich auf einen relativ eingeschränkten inhaltlichen Bereich (insbesondere auf das Unterrichten), viele andere professionelle Fähigkeiten und erzieherische Kompetenzen werden dadurch jedoch aus-

geblendet bzw. nicht berücksichtigt. Weiterhin monieren die Autoren, dass bei jedem zweiten Instrument unklar bleibt, wie die Beurteilungen im Einzelnen durchgeführt werden. Auch sei die testdiagnostische Qualität der Bewertungsinstrumente in keinem Fall solide geprüft gewesen. Strietholt und Terhart (2009, S. 622) konstatieren abschließend, „[…] dass hinsichtlich der Qualität der Beurteilungsprozesse und -instrumente, die in der zweiten Phase der Lehrerbildung eingesetzt werden, ein dringender Entwicklungsbedarf besteht."

Die Sachlage stellt sich dergestalt dar, dass offenbar eine ganze Reihe an Beurteilungsinstrumenten und Kriterienkatalogen – wenn auch teils mit fraglicher Güte (vgl. Strietholt & Terhart, 2009) – existieren und auch grundsätzlich für Referendare/-innen verfügbar sind, diese die Bewertung jedoch nicht unbedingt als nachvollziehbar und einheitlichen Kriterien folgend erleben (vgl. Döbrich & Abs, 2008). Da nach unserer Kenntnis derzeit noch kein Instrument zur Erfassung der wahrgenommenen Bewertungstransparenz im Referendariat vorliegt, deren genauere Betrachtung jedoch relevant erscheint, setzt sich der Beitrag das Ziel, ein solches Instrument vorzustellen und damit die wahrgenommene Bewertungstransparenz in Bezug auf verschiedene Bewertungsanlässe im Referendariat zu operationalisieren. Im Folgenden charakterisieren wir zunächst die wesentlichen Bewertungsanlässe im Referendariat, bevor wir näher auf das pädagogische Leistungsverständnis und auf Merkmale einer transparenten Leistungsbewertung eingehen. Im empirischen Teil der Arbeit werden Studiendesign und Stichprobe dargestellt, bevor die Entwicklung des Instrumentes zur Erfassung der wahrgenommenen Bewertungstransparenz sowie korrelative und deskriptive Befunde beschrieben werden. Eine Zusammenfassung und Diskussion der zentralen Ergebnisse sowie ein Ausblick schließen den Beitrag inhaltlich ab.

2. Hintergrund

2.1 Bewertungsanlässe im Referendariat

Wenngleich die gesetzliche Regulierung der Bildung in die Kulturhoheit der Bundesländer fällt und somit sowohl zwischen den einzelnen Ländern als auch zwischen verschiedenen Schularten innerhalb der Länder Variationen in den einzelnen Prüfungsbestandteilen und/oder -ausgestaltungen vorzufinden sind, so lassen sich doch Parallelen hinsichtlich der existierenden Bewertungsanlässe im Referendariat herausarbeiten (zusammenfassende Übersichten bzgl. länderspezifischer Prüfungsregelungen im Vorbereitungsdienst für das Lehramt finden sich bei KMK, 2012 sowie KMK, 2013; siehe z. B. LPO II für Bayern und LLPA für Baden-Württemberg).

Seminarveranstaltungen, welche sich in Hauptseminar und Fachseminar des Erst- und Zweitfaches unterscheiden lassen, finden während der gesamten Dauer des Vorbereitungsdienstes ein- bis zweimal in der Woche statt und haben eine sowohl ausbildende als auch bewertende Funktion. Im Seminar werden gesammelte Erfahrungen reflektiert und allgemeindidaktische als auch fachdidaktische Kenntnisse vertieft. In den

meisten Schularten wird von den Leiter/-innen des Studienseminars in Zusammenarbeit mit den Schulleitern/-innen der Einsatzschulen sowie den Seminar- und Betreuungslehrkräften ein Gutachten über das Verhalten der Referendare/-innen im Seminar erstellt, welches die Beurteilung dieser hinsichtlich der Aspekte „Unterrichtskompetenz", „erzieherische Kompetenz" sowie „Handlungs- und Sachkompetenz" beinhaltet.

In *Lehrversuchen* bekommen die Referendare/-innen die Möglichkeit, sich in der Planung, Gestaltung und Durchführung von Unterricht zu üben. Dabei arbeiten die Seminarlehrkräfte vor den ersten Unterrichtsversuchen im Rahmen des Fachseminars zunächst gemeinsam mit den Referendaren/-innen die Zielvorstellungen und die Beurteilungskriterien für Unterrichtsplanung und Unterrichtsgestaltung heraus. Im Anschluss an einen Lehrversuch finden i. d. R. Feedback- und daran anschließende Zielvereinbarungsgespräche zwischen dem/der Referendar/-in und der Seminarlehrkraft statt, damit die Referendare/-innen dadurch eine reflektierte und möglichst realistische Selbsteinschätzung bezüglich ihres Ausbildungsstandes entwickeln können und erfahren, in welchen Bereichen sie noch Verbesserungspotenziale aufweisen.

Prüfungslehrproben sind beurteilte Vorführstunden, die in jedem Bundesland vorgeschrieben sind. Zu der Prüfungslehrprobe zählt i. d. R. eine schriftliche Ausarbeitung, welche die Ziele und den Aufbau der durchzuführenden Stunde darlegt und der Prüfungskommission vorzulegen ist. So hat der/die Studienreferendar/-in in der schriftlichen Lehrdarstellung unter anderem didaktisch-methodische Vorüberlegungen bezüglich der Klasse, des Themas und methodischer Entscheidungen darzulegen und einen detaillierten Unterrichtsverlaufsplan anzuführen.

Im Rahmen des *Kolloquiums* hat der/die Referendar/-in vor Beginn der Prüfung Zeit, um sich in eine konkrete und pädagogisch-psychologisch relevante Fallsituation einzuarbeiten und Handlungsmöglichkeiten zu erarbeiten. Auf Grundlage der jeweiligen Fallsituation werden darauf aufbauend vonseiten der prüfenden Seminarlehrkräfte Fragen aus den Gebieten der Pädagogik, Psychologie und beim Lehramt an Sonderschulen auch aus dem Gebiet der Sonderpädagogik gestellt, welche der/die Referendar/-in mündlich beantwortet.

Die *mündliche Prüfung* hat die Prüfung in der jeweiligen Fachdidaktik des Erst- und ggf. Zweitfaches sowie die Themenbereiche Schulrecht bzw. Landesrecht und Staatskunde bzw. Grundfragen der staatsbürgerlichen Bildung zum Inhalt. Die Prüfungsinhalte für die mündliche Prüfung in Schulrecht werden in erster Linie in regelmäßigen Sitzungen mit den jeweiligen Schulleitern im Fachseminar behandelt. Da die mündliche Prüfung zeitlich an das Kolloquium gekoppelt ist bzw. sich diesem i. d. R. unmittelbar anschließt, behandeln wir beide Aspekte in unten beschriebener Fragebogenentwicklung als einen Bewertungsanlass.

Die *Seminararbeit* bzw. schriftliche Hausarbeit umfasst ein Thema aus dem Wissens- und Erfahrungsbereich des/der Lehramtsanwärter/-in. Die Arbeit kann sich auf die Themengebiete Pädagogik/Psychologie oder auf die jeweilige Didaktik des studierten Faches beziehen. Je nach Art des Lehramtes liegt der Fokus auf der Grundschule, den Fächergruppen der Hauptschule, der sonderpädagogischen Fachrichtung oder der beruflichen Fachrichtung (vgl. KMK, 2012; KMK, 2013).

Die Durchführung der Bewertung der einzelnen Prüfungsbestandteile erfolgt zu unterschiedlichen Teilen durch die Leiter/-innen des Studienseminars, die Schulleiter/-innen der Einsatzschulen sowie die Seminarlehrkräfte.

2.2 Pädagogisches Leistungsverständnis und Merkmale einer transparenten Leistungsbewertung

Klafki (1996, S. 228) definiert Leistung als den „Vollzug und das Ergebnis einer Tätigkeit, die mit Anstrengung verbunden, auf die Erlangung eines Zieles gerichtet und auf Gütemaßstäbe und Anforderungen bezogen ist". Das pädagogische Leistungsverständnis nach Klein (2009) beinhaltet u. a., dass Leistung norm- und zweckbezogen ist, damit seitens der Bewerteten erkennbar ist, was anerkannt wird und was erreicht werden soll. Leistung wird weiterhin als produkt- *und* prozessorientiert angesehen, da neben dem Ergebnis auch der Lernprozess in die Bewertung einfließen kann (vgl. Brouër, 2014). Idealerweise ist der-/diejenige, welche/r bewertet wird, mit in den Entscheidungsprozess einzubinden, was als Leistung erbracht und anerkannt werden soll (vgl. Heid, 1992; 2012). Leistungsbewertungsprozesse erfüllen verschiedene Funktionen. So zählt die Bewertung von Leistungen zu den alltäglichen Aufgaben einer jeden (Seminar-)Lehrkraft und hat ihren Ursprung in der notwendigen Auslese aufgrund gesellschaftlicher Zwecke und Ziele, im Sinne einer *Selektionsfunktion.* Eine weitere Funktion der Leistungsbewertung ist die *Sozialisationsfunktion,* indem Handlungsresultate mit gesellschaftlich geltenden Leistungsmaßstäben bewertet werden, was zu einer sukzessiven Adaptation an gesellschaftliche bzw. berufsgruppenspezifische Normen und Ansprüche führen soll. Aus pädagogischer Sicht hat die Leistungsbewertung weiterhin eine *Rückmeldefunktion,* indem Lernende durch sie Feedback bezüglich ihres Leistungstandes und der Leistungsentwicklung erhalten, sowie eine *Anreizfunktion,* indem gute Leistungsbewertungen einen Anreiz für zukünftige Anstrengungen schaffen, wohingegen schlechte Leistungsbewertungen als Ansporn zur Verbesserung dienen sollen (vgl. Jürgens & Sacher, 2000; Oehmig, 2005; Saldern, 1999; Strietholt & Terhart, 2009; Wengert, 2006). Um die Leistung als Beleg des Auseinandersetzungsprozesses mit einem bestimmten Gegenstandsbereich beurteilen zu können, muss einerseits deren Güte bzw. Qualität erfasst werden und andererseits ein Abgleich mit einer vorher festgelegten Norm erfolgen (vgl. Jürgens, 1998; Rheinberg, 2001). Das Ergebnis der Leistungsbeurteilung sollte dem/der Bewerteten mitgeteilt werden, indem Verbesserungspotenziale und Lernfortschritte aufgezeigt werden (vgl. Brouër, 2014; Schrader & Helmke, 2001; Stritzke, Harles & Wenzel, 1982). Die Beurteilenden sind in diesem Kontext stets angehalten, die wesentlichen Gütekriterien der Leistungsbeurteilung (Objektivität, Reliabilität, Validität) einzuhalten, um den Bewertungsprozess möglichst nachvollziehbar, verstehbar und konsistent zu gestalten, was in der pädagogischen Praxis aufgrund von auftretenden Beurteilungsfehlern bzw. -verzerrungen jedoch häufig nicht der Fall ist (vgl. Hartog & Rhodes, 1974; Ingenkamp, 1974; Schrader & Helmke, 2001; Starch & Elliot, 1974).

In der Mehrdimensionalität des pädagogischen Leistungs(bewertungs)verständnisses ist somit bereits die Forderung nach *klaren Grundsätzen* in der Leistungsbewertung begründet. Die wesentlichen Grundlagen von Bewertungstransparenz sind durchschaubare, nachvollziehbare und verständliche Leistungserwartungen und Bewertungskriterien. *Transparente Leistungserwartungen* stehen für klar kommunizierte und verbindliche inhaltliche Ziele, die dem/der Lernenden verbal oder nonverbal vermitteln, welche Leistungen in einer Bewertungssituation zu erbringen sind. Den Lernenden sollte ein an den gültigen Richtlinien oder an Bildungsstandards ausgerichtetes und ihrem Leistungsvermögen angepasstes Lernangebot gemacht werden, das verständlich kommuniziert und zum Gegenstand eines Arbeitsbündnisses zwischen Lehrenden und Lernenden gemacht wird. Weiterhin sollen den Lernenden durch formelle und informelle Leistungskontrollen zeitnah Rückmeldungen zum Lernfortschritt gegeben werden (vgl. Meyer, 2004; Paradies, Wester & Greving, 2005; Schrader & Helmke, 2001).

Eine transparente Leistungsbewertung zeichnet sich weiterhin dadurch aus, dass sich diese an allgemeingültigen und universal akzeptierten Standards und Kriterien orientiert (vgl. Klein, 2009). *Transparente Bewertungskriterien* legen den Lernenden dar, wie die Bewertung der Leistung im Detail erfolgt. In der Bewertungspraxis gibt es zur Transparenzherstellung unterschiedliche Möglichkeiten. Beispielsweise können Bewertungsbögen, Kriterienkataloge oder Beurteilungsraster mit Punkteverteilungsschlüsseln eingesetzt werden, um die Bewertung nachvollziehbar und möglichst objektiv durchführen zu können (vgl. Strietholt & Terhart, 2009). Offen und verständlich kommunizierte sowie nachvollziehbare Bewertungskriterien können den Lernenden Einblicke in Qualität und Quantität ihrer individuellen Verbesserungspotenziale geben und somit die persönliche Lernentwicklung fördern. Weiterhin tragen klar formulierte Bewertungskriterien dazu bei, dass die gleichen Fehler von verschiedenen Personen nach den gleichen Kriterien bewertet werden. Idealerweise sollten die Lernenden in die Erstellung der Bewertungskriterien einbezogen werden, um die Nachvollziehbarkeit und Akzeptanz durch die Bewerteten zu erhöhen. Außerdem sollten die Erwartungen und Bewertungskriterien vor einer Prüfung offen kommuniziert werden (vgl. Bohl, 2003; 2004; Klein, 2009; Paradies, Wester & Greving, 2005).

Ein nachvollziehbarer *Rückmeldeprozess* sowie eine *positive Fehlerkultur,* welche u. a. das Lernen aus Fehlern in einem geschützten Rahmen ermöglicht (vgl. Oser, Hascher & Spychiger, 1999), sind ebenfalls unumgänglich für eine transparente Leistungsbewertung. Im Rückmeldeprozess sollten vom Feedbackgeber Informationen an den Feedbacknehmer gegeben werden, auf deren Grundlage individuelle Bestätigungs-, Anpassungs- oder Umstrukturierungsprozesse angestoßen werden können (vgl. Winne & Butler, 1994). Feedback sollte sich stets direkt auf den konkreten und individuellen Lernprozess beziehen, um vorhandene Schwächen erkennen und auszugleichen sowie vorhandene Stärken ausbauen zu können (vgl. Hattie & Timperley, 2007). Eine Möglichkeit für eine angemessene Rückmeldekultur ist beispielsweise ein Bewertungsgespräch mit anschließender Beratung über die nächsten Lernziele oder ein Lernentwicklungsbericht, in welchem die individuelle Lernsituation inklusive persönlicher Lernfortschritte und Lerndefizite verständlich und nachvollziehbar dargestellt werden.

Vergleichbare Möglichkeiten bieten Diagnosebögen, die Informationen für die individuelle Lernentwicklung anhand aktueller Kompetenzen und Defizite beinhalten (vgl. Bohl, 2003; 2004).

3. Fragestellung und Annahmen

Vor dem Hintergrund der in der Literatur vorzufindenden Merkmale einer transparenten Leistungsbewertung gehen wir den Fragen nach, inwieweit sich die identifizierten Dimensionen in Bezug auf unterschiedliche Bewertungsanlässe im Referendariat empirisch abbilden lassen und wie transparent die Bewertung seitens der Befragten wahrgenommen wird.

Wir nehmen an, dass sich transparente Leistungserwartungen und Bewertungskriterien einerseits und konstruktives Feedback andererseits als zentrale, jedoch nicht voneinander unabhängige, Dimensionen einer transparenten Leistungsbewertung empirisch abbilden lassen. In Bezug auf die erwartete Ausprägung der wahrgenommenen Bewertungstransparenz gehen wir davon aus, dass diese im Mittel eher mäßig eingeschätzt wird (vgl. Döbrich & Abs, 2008; Strietholt & Terhart, 2009).

Hinsichtlich der betrachteten Außenkriterien nehmen wir an, dass die Dimensionen einer transparenten Leistungsbewertung in positivem Zusammenhang mit der Zufriedenheit mit der Ausbildung im Referendariat stehen (vgl. Drüge, Schleider & Rosati, 2014) sowie in negativem Zusammenhang mit der Belastung bzgl. der Ungewissheit über die Benotung im Referendariat (vgl. Kaluza, 2015; Rost & Schermer, 2010).

4. Methode

4.1 Studiendesign und Stichprobe

Um der Frage nach der wahrgenommenen Bewertungstransparenz im Referendariat nachzugehen und in diesem Zusammenhang ein Instrument zur Erfassung des genannten Konstrukts zu entwickeln, wurde eine anonyme und onlinebasierte Datenerhebung durchgeführt. Da es sich bei Lehramtsanwärtern/-innen sowie fertig ausgebildeten Lehrkräften um eine eher schwer zu erreichende Stichprobe handelt, haben wir uns bei der Datenerhebung für das Schneeballverfahren entschieden (vgl. Döring & Bortz, 2016). Die Einladung zur Befragung wurde über universitäre Mailverteiler an eingeschriebene und ehemalige Studierende des Lehramts versandt sowie in einschlägigen Foren für Lehramtsanwärter/-innen veröffentlicht. Die angeschriebenen Personen wurden darüber hinaus auch gebeten, die Anfrage an ihnen bekannte angehende, aktive und ehemalige Lehrkräfte weiterzuleiten.

Insgesamt riefen 263 Personen den bereitgestellten Link zum Fragebogen auf. Ca. 100 Personen brachen die Bearbeitung bereits vor den relevanten Items ab, womit letztendlich 164 verwertbare Fragebögen vorlagen. Zwölf Personen, deren Referendariat

zum Zeitpunkt der Befragung mehr als zehn Jahre (*M* = 28.17, *SD* = 10.36, *Min.* = 12, *Max.* = 44) zurücklag, wurden aus der Stichprobe herausgenommen, da deren Einschätzungen aufgrund der sehr hohen Retrospektivität und auch aufgrund zwischenzeitlicher struktureller Reformen der zweiten Phase der Lehrer/-innenbildung als nicht vergleichbar mit den Einschätzungen der restlichen Stichprobe angenommen wurden.

Die finale Stichprobe setzt sich somit aus insgesamt 152 Personen (120 weiblich, 32 männlich) mit einem durchschnittlichen Alter von 30.56 Jahren (*SD* = 4.14, *Min.* = 24, *Max.* = 56) zusammen. Die Befragten absolvierten ihr Referendariat in den folgenden Bundesländern: Bayern (*n* = 131, 86.2 %), Baden-Württemberg (*n* = 10, 6.6 %), Nordrhein-Westfalen (*n* = 5, 3.3 %), Hessen (*n* = 3, 2.0 %), Niedersachsen (*n* = 2, 1.3 %) und Berlin (*n* = 1, .7 %). Zum Zeitpunkt der Befragung befanden sich insgesamt 43 Personen im Referendariat, wobei sich die Befragten jeweils ca. hälftig auf das erste und zweite Referendariatsjahr verteilen. Von den restlichen 109 Personen hatten 107 das Referendariat zum Zeitpunkt der Befragung bereits komplett und erfolgreich abgeschlossen. Im Mittel lag ihr Referendariat 2.52 Jahre (SD = 2.05) zurück. Zwei Personen in der Stichprobe haben das Referendariat unter- bzw. abgebrochen. Im ersten Jahr des Referendariats waren die Befragten mehrheitlich an Gymnasien (46.2 %) tätig, gefolgt von beruflichen Schulen (Berufsschule, Fach-/Berufsoberschule, Berufsfachschule, Wirtschaftsschule) (23.7 %), Realschulen (16.6 %), Mittelschulen (3.6 %) und Grundschulen (5.3 %). In 4.7 % der Fälle waren Personen im ersten Jahr an anderen Schultypen (z. B. Gesamtschulen, Förderschulen) tätig. Die Verteilung spiegelt sich auch im zweiten Referendariatsjahr wider: 47.8 % der Befragten waren an einem Gymnasium, 19.9 % an einer beruflichen Schule, 16.8 % an einer Realschule, 6.2 % an einer Grundschule und 3.1 % an einer Mittelschule. In 6.2 % der Fälle waren andere Schultypen involviert (z. B. Gesamtschulen, Förderschulen).

4.2 Statistische Analysen

Zunächst wurde die Faktorenstruktur für jeden der fünf Bewertungsanlässe mittels explorativer Faktorenanalyse (Hauptachsenanalyse mit Promax Rotation, Eigenwert > 1) unter Verwendung der Software SPSS 25® (IBM®, Chicago, USA) geprüft. Die konfirmatorische Prüfung der Faktorenstruktur erfolgte mit der Software Mplus7 (vgl. Muthén & Muthén, 1998–2010) unter Rückbezug auf die gängigen Gütekriterien. Weiterhin wurden die deskriptiven Ergebnisse für die Gesamtskalenmittelwerte sowie für die Subskalen „Transparente Leistungserwartungen und Bewertungskriterien“ und „Konstruktives Feedback“ und die Skaleninterkorrelationen und Korrelationen mit den Außenkriterien und soziodemografischen Merkmalen (via Pearson Produkt-Moment Korrelationen) mit der Software SPSS 25® (IBM®, Chicago, USA) berechnet.

5. Entwicklung eines Instrumentes zur Erfassung der wahrgenommenen Bewertungstransparenz im Referendariat

5.1 Vorgehen bei der Itementwicklung

Ausgehend von den herausgearbeiteten Merkmalen einer transparenten Leistungsbewertung (siehe Abschnitt 2.2) wurden für jeden Bewertungsanlass im Referendariat (siehe Abschnitt 2.1) entsprechende Items formuliert. Die Items zu den *transparenten Leistungserwartungen und Bewertungskriterien* beinhalten Aussagen zur wahrgenommenen Durchschaubarkeit, Nachvollziehbarkeit, Verständlichkeit, Verbindlichkeit sowie zur offenen Kommunikation der formulierten Leistungserwartungen und Bewertungskriterien. Die Items zum *konstruktiven Feedback* beziehen sich auf einen nachvollziehbaren und verständlichen Rückmelde- bzw. Feedbackprozess, durch den die eigenen Stärken und Schwächen aufgezeigt werden und konstruktiv und verantwortungsvoll mit diesen umgegangen wird.

Der Prozess der Itemformulierung durchlief verschiedene Stufen. Zunächst wurden insgesamt 113 Rohitems formuliert und in ihrer Formulierung sukzessive verfeinert. Im Anschluss wurden die Items von zwei Experten/-innen bewertet, die beide voll ausgebildete Lehrkräfte sind und zum Zeitpunkt der Befragung mit der Ausbildung von Referendare/-innen betraut waren. Auf Basis der Experten/-innenrückmeldung wurden die Formulierungen der Items nochmals optimiert. Die Items wurden auf einer sechsstufigen Likert-Skala bewertet von 1 = „trifft nicht zu" bis 6 = „trifft zu" und blockweise entsprechend der jeweiligen Bewertungsanlässe abgefragt.[1]

Abgesehen von den Items zur wahrgenommenen Bewertungstransparenz wurden die *Zufriedenheit mit der Ausbildung im Referendariat* (1 = „sehr unzufrieden" bis 5 = „sehr zufrieden"; $M = 2.91$, $SD = 1.12$, $Min. = 1$, $Max. = 5$) sowie die wahrgenommene *Belastung im Zusammenhang mit der Ungewissheit über die Benotung im Referendariat* erfragt (1 = „überhaupt nicht belastend" bis 4 = „sehr belastend"; $M = 3.53$, $SD = .69$, $Min. = 1$, $Max. = 4$; jeweils Eigenentwicklungen) und als Außenkriterien korrelativ mit den Skalen einer transparenten Leistungsbewertung in Verbindung gesetzt.

5.2 Itemauswahl und deskriptive Itemkennwerte

Zunächst wurden die finalen 40 Items (je 20 Items für den Faktor „Transparente Leistungserwartungen und Bewertungskriterien" und für den Faktor „Konstruktives Feedback") auf Grundlage von deskriptiven Itemkennwerten wie Mittelwert, Standardabweichung und Trennschärfe, sowie der Ladungen bei einer explorativen Faktorenanalyse ausgewählt. Für jeden der fünf Bewertungsanlässe ergeben sich die Faktoren „Transparente Leistungserwartungen und Bewertungskriterien" (TLB) und „Konstruktives

1 Der vollständige ‚Fragebogen zur Erfassung der wahrgenommenen Bewertungstransparenz im Referendariat' kann bei Tobias Kärner (tobias.kaerner@uni-konstanz.de) erfragt werden.

Feedback“ (KF), die jeweils mit vier Items gemessen werden. Zusammengenommen ergeben die Subskalen TLB und KF die Gesamtskala „Bewertungstransparenz“ für jeden der fünf Bewertungsanlässe.

Mittelwert, Standardabweichung, Trennschärfe und Itemschwierigkeit der einzelnen Items sind in Tabelle 1 aufgeführt. Die Mittelwerte der Items bewegen sich im Bereich von 2.49 (wobei 2 = „trifft überwiegend nicht zu“) bis 4.29 (wobei 4 = „trifft eher zu“). In Ergänzung zu den Mittelwerten fallen die teils großen Standardabweichungen der einzelnen Items auf, welche sich im Bereich von 1.29 bis 1.87 bewegen. Die Trennschärfen der Items erweisen sich als angemessen und liegen im Bereich zwischen .61 und .90. Die Itemschwierigkeiten rangieren im mittleren Bereich, mit Werten zwischen .30 und .66. Die Skalenreliabilitäten, gemessen an Cronbachs Alpha, liegen bei allen TLB- und KF-Skalen im Bereich zwischen .85 und .95. Die Cronbachs Alpha Werte für die Bewertungstransparenz (Gesamtskala) nach Bewertungsanlass liegen im Bereich zwischen .91 und .95 und sind somit ebenfalls als adäquat zu bewerten.

5.3 Konfirmatorische Prüfung der Faktorenstruktur

Für jeden Bewertungsanlass wurde die Faktorenstruktur konfirmatorisch überprüft. Abbildung 1 zeigt die für jeden Bewertungsanlass angenommene Faktorenstruktur. Da sich literaturbasierte Hinweise darauf ergeben, dass die Faktoren TLB und KF zwar als zentrale, jedoch nicht voneinander unabhängige Dimensionen einer als transparent wahrgenommenen Leistungsbewertung anzunehmen sind (vgl. Bohl, 2003; 2004; Klein, 2009; Meyer, 2004; Paradies, Wester & Greving, 2005; Schrader & Helmke, 2001), ist davon auszugehen, dass die beiden Faktoren einem übergeordneten Faktor „Bewertungstransparenz“ zugrunde liegen.

Die Fit-Indizes für die einzelnen Modelle, gemessen an CMIN/DF, RMSEA, CFI, TLI, und SRMR, weisen auf eine gute bis sehr gute Anpassungsgüte hin (Tab. 2).

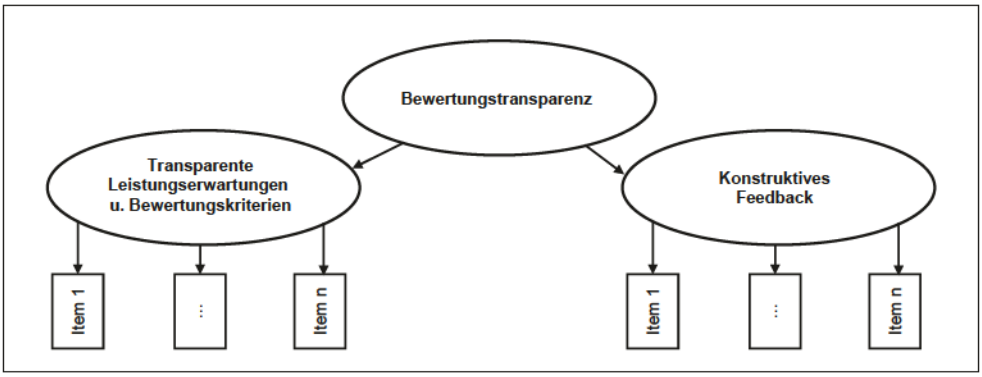

Abb. 1: Messmodell „Bewertungstransparenz“

Bewertungsanlass	Item	Itemformulierung	*M*	*SD*	r_{it}	P_i
Seminarveranstaltungen (SE), α = .91						
Transparente Leistungserwartungen und Bewertungskriterien N = 152 α = .87	SE_01	In den Seminarveranstaltungen waren die Bewertungskriterien der einzelnen Prüfungen, die ich im Referendariat durchlaufen musste, für mich verständlich und nachvollziehbar formuliert.	3.45	1.38	.73	.49
	SE_02	Mir waren die inhaltlichen Ziele der Seminarveranstaltungen klar.	3.91	1.35	.73	.58
	SE_03	In den Seminarveranstaltungen wurde für mich verständlich und nachvollziehbar formuliert, welche Leistungen ich erbringen sollte.	3.55	1.29	.71	.51
	SE_04	Mir war klar, zu welchem Ziel bestimmte Anforderungen in den Seminarveranstaltungen führen sollten.	3.49	1.36	.74	.50
Konstruktives Feedback N = 152 α = .85	SE_05	In den Seminarveranstaltungen erhielt ich konstruktives Feedback zu meinem Ausbildungsstand.	3.07	1.32	.71	.41
	SE_06	In den Seminarveranstaltungen erhielt ich für mich verständliches und nachvollziehbares Feedback zu meinem Ausbildungsstand.	3.22	1.32	.77	.44
	SE_07	Durch die Seminarveranstaltungen wurden mir meine eigenen Stärken und Schwächen aufgezeigt.	3.20	1.29	.69	.44
	SE_08	In den Seminarveranstaltungen wurde ich individuell gefördert.	2.49	1.38	.61	.30
Unterrichtsbesuche (UB), α = .94						
Transparente Leistungserwartungen und Bewertungskriterien N = 134 α = .91	UB_01	Erwartungen an mich während des Unterrichtsbesuches wurden offen kommuniziert.	3.55	1.56	.83	.51
	UB_02	Die Bewertungskriterien des Unterrichtsbesuches waren für mich verständlich und nachvollziehbar.	3.48	1.59	.84	.50
	UB_03	Mir wurde klar aufgezeigt, welche Kompetenzen ich in den Unterrichtsbesuchen zeigen sollte.	3.22	1.61	.79	.44
	UB_04	Die Bewertungskriterien des Unterrichtsbesuches wurden mir vorab erläutert.	2.85	1.57	.72	.37
Konstruktives Feedback N = 134 α = .95	UB_05	Ich erhielt für meinen gehaltenen Unterricht konstruktives Feedback.	4.29	1.47	.88	.66
	UB_06	Mit meinen Fehlern während des Unterrichtsbesuches wurde konstruktiv umgegangen.	3.97	1.53	.90	.59
	UB_07	Das gegebene Feedback für den gehaltenen Unterricht war für mich verständlich und nachvollziehbar.	3.96	1.54	.88	.59
	UB_08	Das Feedback des Unterrichtsbesuches half mir, meine eigenen Schwächen und Stärken zu erkennen.	3.99	1.50	.86	.60

Bewertungsanlass	Item	Itemformulierung	*M*	*SD*	r_{it}	P_i
Prüfungslehrproben (PL), α = .95						
Transparente Leistungserwartungen und Bewertungskriterien N = 108 α = .93	PL_01	Die Bewertungskriterien für die Prüfungslehrprobe waren für mich verständlich und nachvollziehbar.	3.31	1.74	.87	.46
	PL_02	Mir wurde klar aufgezeigt, welche Kompetenzen ich in der Prüfungslehrprobe umsetzen sollte.	3.50	1.63	.79	.50
	PL_03	Die Bewertungskriterien für die Prüfungslehrprobe wurden vorab erläutert.	3.51	1.67	.82	.50
	PL_04	Die Bewertungskriterien für die Prüfungslehrprobe wurden mir offen kommuniziert.	3.41	1.64	.87	.48
Konstruktives Feedback N = 107 α = .94	PL_05	Ich erhielt für meinen gehaltenen Unterricht in der Prüfungslehrprobe konstruktives Feedback.	3.82	1.70	.88	.56
	PL_06	Mit meinen Fehlern aus der Prüfungslehrprobe wurde konstruktiv umgegangen.	3.64	1.67	.83	.53
	PL_07	Das gegebene Feedback zu meiner Prüfungslehrprobe war für mich verständlich und nachvollziehbar.	3.33	1.70	.90	.47
	PL_08	Das Feedback der Prüfungslehrprobe half mir meine eigenen Schwächen und Stärken zu erkennen.	3.33	1.75	.79	.47
Mündliche Prüfung/Kolloquium (MK), α = .95						
Transparente Leistungserwartungen und Bewertungskriterien N = 97 α = .91	MK_01	Die an mich gestellten Erwartungen in der mündlichen Prüfung/dem Kolloquium wurden offen kommuniziert.	3.77	1.78	.85	.55
	MK_02	Die Bewertungskriterien der mündlichen Prüfung/des Kolloquiums waren für mich verständlich und nachvollziehbar.	3.58	1.73	.85	.52
	MK_03	Mir wurde klar aufgezeigt, welche Inhalte ich in der mündlichen Prüfung/dem Kolloquium beherrschen musste.	3.55	1.83	.82	.51
	MK_04	Die Bewertungskriterien der mündlichen Prüfung/des Kolloquiums wurden vorab erläutert.	4.05	1.63	.68	.61
Konstruktives Feedback N = 93 α = .92	MK_05	Ich erhielt für meine mündliche Prüfung/mein Kolloquium konstruktives Feedback.	3.44	1.73	.83	.49
	MK_06	Das gegebene Feedback bezüglich der mündlichen Prüfung/des Kolloquiums war für mich verständlich und nachvollziehbar.	3.65	1.74	.83	.53
	MK_07	Das Feedback bezüglich der mündlichen Prüfung/des Kolloquiums half mir meine eigenen Schwächen und Stärken zu erkennen.	3.01	1.79	.85	.40
	MK_08	Mögliche Maßnahmen zur Verbesserung meiner Leistung der mündlichen Prüfung/des Kolloquiums wurden mir unterbreitet.	2.55	1.80	.76	.31

Bewertungsanlass	Item	Itemformulierung	*M*	*SD*	r_{it}	P_i
Seminararbeit (SA), α = .91						
Transparente Leistungserwartungen und Bewertungskriterien N = 111 α = .89	SA_01	Die Rahmenbedingungen der Seminararbeit (z. B. Zitationsstil, Schrift, Umfang usw.) wurden klar kommuniziert.	4.04	1.87	.61	.61
	SA_02	Die Bewertungskriterien der Seminararbeit wurden vorab erläutert.	3.35	1.86	.81	.47
	SA_03	Es wurde für mich verständlich und nachvollziehbar formuliert, welche Leistungen ich in der Seminararbeit erbringen sollte.	3.41	1.70	.84	.48
	SA_04	Die Bewertungskriterien der Seminararbeit waren mir zu Beginn offen kommuniziert worden.	3.38	1.80	.81	.48
Konstruktives Feedback N = 93 α = .90	SA_05	Ich erhielt für meine Seminararbeit konstruktives Feedback.	3.03	1.75	.86	.41
	SA_06	Das gegebene Feedback für meine Seminararbeit war für mich verständlich und nachvollziehbar.	3.03	1.85	.83	.41
	SA_07	Das Feedback für meine Seminararbeit half mir meine eigenen Schwächen und Stärken zu erkennen.	2.59	1.69	.86	.32
	SA_08	Das Feedback für meine Seminararbeit erhielt ich zeitnah.	2.83	1.77	.61	.37

Anmerkungen: r_{it} = Itemtrennschärfe; P_i = Itemschwierigkeit; sechsstufige Likert-Skala: 1 = „trifft nicht zu", 2 = „trifft überwiegend nicht zu", 3 = „trifft eher nicht zu", 4 = „trifft eher zu", 5 = „trifft überwiegend zu", 6 = „trifft zu". Die unterschiedlichen Stichprobengrößen sind u. a. darauf zurückzuführen, dass nicht alle Teilnehmer/-innen zum Zeitpunkt der Befragung auch bereits alle der abgefragten Bewertungsanlässe durchlaufen hatten. Im Falle, dass sie eine Aussage nicht beurteilen konnten, weil sie den betreffenden Bewertungsanlass noch nicht absolviert hatten bzw. dieser nicht im Ausbildungsprogramm vorgesehen war, hatten die Teilnehmer/-innen die Möglichkeit, die Option „noch nicht absolviert" auszuwählen.

Tab. 1: Itemkennwerte und Reliabilitäten der Skalen und Subskalen

	N	χ^2	*p*	*df*	CMIN/DF	RMSEA	CFI	TLI	SRMR
Seminarveranstaltungen	152	27.82	.087	19	1.46	.06	.98	.97	.04
Unterrichtsbesuche	136	35.44	.012	19	1.87	.08	.98	.97	.03
Prüfungslehrprobe	108	29.56	.058	19	1.56	.07	.98	.97	.03
Mündliche Prüfung/Kolloquium	98	30.23	.049	19	1.59	.08	.98	.97	.03
Seminararbeit	113	22.08	.280	19	1.16	.04	.99	.99	.02

Tab. 2: Modelltests und Fit-Indizes

5.4 Skaleninterkorrelationen und Korrelationen mit Außenkriterien sowie soziodemographischen Angaben

Bezüglich der Korrelationen zwischen den Subskalen der einzelnen Bewertungsanlässe, zeigen sich Pearson Produkt-Moment Korrelationen im Bereich von $r = .36$ bis $r = .71$ (TLB) und von $r = .16$ bis $r = .66$ (KF) (Tab. 3). Innerhalb der verschiedenen Bewertungsanlässe korrelieren die Subskalen TLB und KF im Bereich von $r = .59$ bis $r = .82$ miteinander und die Gesamtskalen korrelieren im Bereich von $r = .27$ bis $r = .72$ miteinander.

Hinsichtlich des Außenkriteriums der *Zufriedenheit mit der Ausbildung im Referendariat* zeigt sich, dass diese – bis auf die Skalen „Mündliche Prüfung/Kolloquium" (Gesamtskala, KF, TLB) sowie „Seminararbeit" (KF) – in signifikant positivem Zusammenhang mit der wahrgenommenen Bewertungstransparenz steht. Bei der *Belastung im Zusammenhang mit der Ungewissheit über die Benotung im Referendariat* zeigt sich ein entgegengesetztes Bild. Diese steht mit den Transparenzkriterien durchgängig in negativem Zusammenhang, wobei dieser auch hier für die Skalen „Mündliche Prüfung/Kolloquium" (Gesamtskala, KF, TLB) sowie „Seminararbeit" (KF) nicht signifikant ausfällt.

Hinsichtlich der Zusammenhänge mit soziodemographischen Angaben zeigen die korrelativen Auswertungen, dass ältere Personen die Bewertungstransparenz hinsichtlich der Unterrichtsbesuche (Gesamtskala, $r = -.19$; KF, $r = -.22$) sowie hinsichtlich der Prüfungslehrproben (Gesamtskala, $r = -.20$; TLB, $r = -.22$) signifikant geringer bewerten. Beim Geschlecht zeigen sich keine überzufälligen Zusammenhänge mit der Bewertungstransparenz. Personen, die sich zum Zeitpunkt der Befragung im Referendariat befanden, bewerteten die Bewertungstransparenz bzgl. der Seminarveranstaltungen (KF, $r = -.16$), der Unterrichtsbesuche (Gesamtskala, $r = -.21$; KF, $r = -.29$) sowie der Prüfungslehrproben (Gesamtskala, $r = -.22$; KF, $r = -.29$) signifikant höher als Personen, die das Referendariat bereits abgeschlossen hatten. Dagegen schätzen letztgenannte Personen die Bewertungstransparenz bei der mündlichen Prüfung und beim Kolloquium (Gesamtskala, $r = .21$; TLB, $r = .21$; KF, $r = .23$) im Vergleich zu erstgenannten Personen signifikant höher ein. Der Zeitraum nach Abschluss des Referendariats in Jahren steht mit keinen der erfassten Transparenzkriterien in einem bedeutsamen Zusammenhang.

5.5 Deskriptive Ergebnisse für die Gesamtskalenmittelwerte sowie für die Subskalen

Abbildung 2 veranschaulicht die deskriptiven Ergebnisse für die Gesamtskalenmittelwerte sowie für die Skalen TLB und KF. Auffällig sind insbesondere die relativ geringen Ausprägungen beim konstruktiven Feedback bei der Seminararbeit ($M = 2.84$) sowie bei den Seminarveranstaltungen ($M = 3.00$; wobei 3 = „trifft eher nicht zu"). Die deskriptiv höchste Ausprägung zeigt sich beim konstruktiven Feedback, welches im Rahmen von Unterrichtsbesuchen gegeben wird bzw. wurde ($M = 4.08$; wobei 4 = „trifft

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
1. SE ges.																				
2. SE (TLB)	.93***																			
3. SE (KF)	.93***	.73***																		
4. UB ges.	.67***	.60***	.64***																	
5. UB (TLB)	.68***	.62***	.65***	.94***																
6. UB (KF)	.58***	.51***	.57***	.94***	.78***															
7. PL ges.	.63***	.56***	.62***	.72***	.70***	.68***														
8. PL (TLB)	.69***	.63***	.66***	.69***	.71***	.60***	.93***													
9. PL (KF)	.49***	.41***	.51***	.66***	.59***	.66***	.94***	.75***												
10. MK ges.	.46***	.46***	.40***	.27**	.34***	.16	.31**	.35***	.23*											
11. MK (TLB)	.53***	.52***	.46***	.28**	.36***	.15	.30**	.39***	.17	.95***										
12. MK (KF)	.36***	.38***	.30**	.23*	.29**	.16	.28**	.27**	.26**	.95***	.82***									
13. SA ges.	.57***	.53***	.52***	.42***	.45***	.33***	.48***	.52***	.38***	.35***	.38***	.30**								
14. SA (TLB)	.58***	.53***	.55***	.46***	.50***	.36***	.46***	.52***	.35***	.33***	.37***	.27**	.90***							
15. SA (KF)	.42***	.41***	.37***	.25**	.31***	.17	.37***	.39***	.30**	.32**	.31**	.30**	.90***	.59***						
16. Alter	−.13	−.12	−.13	−.19*	−.14	−.22*	−.20*	−.22*	−.15	.14	.11	.17	−.05	−.02	−.05					
17. Geschl.[a]	.03	−.01	.06	.08	.10	.07	−.01	.05	−.07	−.01	−.01	−.02	−.06	.02	−.17	.14				
18. Ref.[b]	−.12	−.05	−.16*	−.21*	−.09	−.29***	−.22*	−.11	−.29**	.21*	.21*	.23*	−.15	−.15	.06	.32***	−.03			
19. Zeitr. Ref.[c]	.11	.09	.12	.12	.11	.12	.07	.02	.11	.14	.11	.16	.08	.04	.13	.44***	.02	–		
20. Zufriedenh.	.51***	.44***	.49***	.65***	.55***	.67***	.68***	.58***	.69***	.12	.06	.16	.25**	.29**	.15	−.20*	−.09	−.27**	.24*	
21. Belastung	−.30***	−.26***	−.30***	−.29***	−.26**	−.28***	−.39***	−.35***	−.38***	−.15	−.14	−.16	−.25**	−.31***	−.12	.01	−.10	.04	−.28**	−.33***

Anmerkungen. Abkürzungen: SE = Seminarveranstaltungen, UB = Unterrichtsbesuche, PL = Prüfungslehrproben, MK = Mündliche Prüfung/Kolloquium, SA = Seminararbeit, TLB = Transparente Leistungserwartungen und Bewertungskriterien, KF = Konstruktives Feedback; [a] Geschlecht: 0 = weiblich, 1 = männlich; [b] Zum Zeitpunkt der Befragung im Referendariat: 0 = Ja, 1 = Nein; [c] Zeitraum nach Abschluss des Referendariats in Jahren; *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$; $86 \leq n \leq 152$

Tab. 3: Skaleninterkorrelationen und Korrelationen mit Außenkriterien

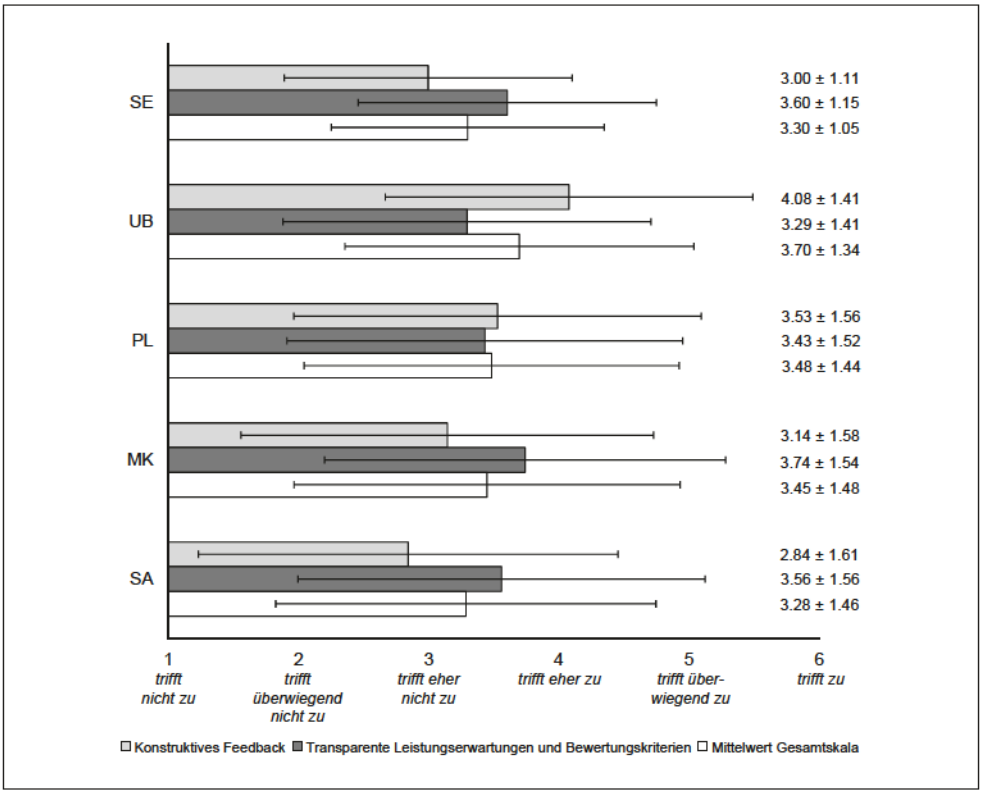

Abb. 2: Deskriptive Ergebnisse für die Gesamtskalenmittelwerte sowie für die Subskalen TLB und KF (jeweils M ± SD; SE = Seminarveranstaltungen, UB = Unterrichtsbesuche, PL = Prüfungslehrproben, MK = Mündliche Prüfung/Kolloquium, SA = Seminararbeit)

eher zu"). Sämtliche weitere Mittelwerte für die Gesamtskalen sowie für die beiden Subskalen bewegen sich im mittleren Messbereich, was weder auf eine besonders gute, noch besonders schlechte wahrgenommene Bewertungstransparenz hinweist. Weiterhin fallen die hohen Standardabweichungen auf, welche sich im Bereich von 1.05 bis 1.61 bewegen.

6. Diskussion und Ausblick

Ausgehend von berechtigten Zweifeln an einer durchgängig transparenten Bewertungspraxis in der zweiten Phase der Lehrer/-innenbildung einerseits (vgl. Döbrich & Abs, 2008; Strietholt & Terhart, 2009) und der sehr geringen Kenntnis über die seitens angehender Lehrpersonen wahrgenommene Bewertungstransparenz andererseits, war es unser Ziel, ein Instrument zur Erfassung der wahrgenommenen Bewertungstransparenz zu entwickeln. Auf Grundlage existierender Literatur zur pädagogischen Leistungs-

bewertung definierten wir zunächst (1) nachvollziehbare sowie vor einer Prüfungssituation offen kommunizierte Leistungserwartungen und Bewertungskriterien und (2) einen konstruktiven Rückmeldeprozess als zentrale Dimensionen einer transparenten Leistungsbewertung (vgl. Klein, 2009; Oser, Hascher & Spychiger, 1999; Meyer, 2004; Paradies, Wester & Greving, 2005). Die literaturbasierten Überlegungen zu den Merkmalen einer transparenten Leistungsbewertung in Kombination mit verschiedenen Bewertungsanlässen im Referendariat dienten als Grundlage für die Entwicklung eines Instrumentes zur Erfassung der wahrgenommenen Bewertungstransparenz. Die jeweils aus acht Items bestehenden Gesamtskalen für jeden der fünf Bewertungsanlässe weisen gute bis sehr gute Itemkennwerte auf und auch die Reliabilitäten sind als gut bis sehr gut einzuschätzen. Die aus der Literatur abgeleitete Struktur aus den zwei Subskalen „Transparente Leistungserwartungen und Bewertungskriterien" und „konstruktives Feedback" konnte in der konfirmatorischen Faktorenstruktur für jeden Bewertungsanlass entsprechend unserer Annahme mit einer guten bis sehr guten Modellpassung abgebildet werden (Tab. 1, Tab. 2). Hinsichtlich der Skaleninterkorrelationen zeigt sich, dass die Subskalen TLB und KF zwischen den einzelnen Bewertungsanlässen im Mittel geringer miteinander korrelieren ($r = .15$ bis $r = .66$) als innerhalb der Bewertungsanlässe ($r = .59$ bis $r = .82$; Tab. 3), was auf eine Spezifität der wahrgenommenen Bewertungstransparenz bezogen auf die verschiedenen Bewertungsanlässe hinweist. Die Größenunterschiede der Korrelationen zwischen den Gesamtskalen könnten darin begründet sein, dass sich manche Bewertungsanlässe hinsichtlich Inhalten und Modalitäten (z. B. gleiche Prüfungspersonen, eine vergleichbare Kommunikation der Bewertungskriterien in bestimmten Prüfungskontexten oder ähnliche Prüfungsstrukturen) ähnlicher sind als andere (z. B. Unterrichtsbesuche und Prüfungslehrproben, $r = .72$ *vs.* Unterrichtsbesuche und Mündliche Prüfung/Kolloquium, $r = .27$).

Die positiven Korrelationen zwischen den (Sub-)Skalen und der Zufriedenheit mit der Ausbildung im Referendariat und die negativen Zusammenhänge mit der Belastung in Zusammenhang mit der Ungewissheit über die Benotung im Referendariat (Tab. 3) zeigen sich als erwartungskonform und geben erste Hinweise auf die Konstruktvalidität, wenngleich aufgrund der geringen Anzahl berücksichtigter Außenkriterien sowie deren eingeschränkter Operationalisierung (jeweils mittels eines Items) weiterführende Studien zur Konstruktvalidierung notwendig erscheinen. Aus inhaltlichen Gesichtspunkten kann die wahrgenommene Bewertungstransparenz als lohnender Forschungsgegenstand im Rahmen der Lehrer/-innenausbildung angesehen werden. Gerade vor dem Hintergrund, da Lehramtsanwärter/-innen besonderen berufsbedingten Belastungen ausgesetzt sind (vgl. Dietrich, 2018; Drüge, Schleider & Rosati, 2014; Klusmann et al., 2012; Košinár, 2014; 2018; Warwas, Neubauer & Panzer, 2016), kann eine Förderung der Bewertungstransparenz möglicherweise sowohl einen Beitrag zur Erhöhung bzw. Aufrechterhaltung der Arbeits- und Ausbildungszufriedenheit leisten als auch zu einer Belastungsreduktion führen.

Jedoch lässt das gewählte (korrelative) Untersuchungsdesign keine Aussagen hinsichtlich den jeweiligen Wirkungsrichtungen zu. D. h., es kann sein, dass als transparent wahrgenommene Bewertungssituationen zu einer höheren Zufriedenheit sowie zu einer

geringeren Belastung führen; es kann aber auch sein, dass geringe Zufriedenheit in der Ausbildung sowie hohe Belastung dazu führen, dass die Ausbildung als weniger transparent wahrgenommen wird. Um entsprechende Wirkungsrichtungen zu identifizieren und zu verifizieren, sind demnach zumindest längsschnittliche Untersuchungsdesigns erforderlich. In Bezug auf die erhobenen soziodemographischen Merkmale zeigen sich für die entwickelten Skalen keine überzufälligen Variationen in Abhängigkeit des Geschlechts oder des Zeitraums nach Abschluss des Referendariats. Jedoch zeigen die korrelativen Analysen kleinere Unterschiede in der Wahrnehmung der Bewertungstransparenz in Abhängigkeit davon, ob sich die Befragten zum Zeitpunkt der Erhebung im Referendariat befanden oder ob sie dieses bereits abgeschlossen hatten. Auch bewerten ältere Personen die Bewertungstransparenz hinsichtlich bestimmter Facetten signifikant geringer als jüngere Personen. Die diesbezüglichen, zeitbezogenen Wahrnehmungsunterschiede sind möglicherweise auf den individuellen Grad an Retrospektivität zurückzuführen. In diesem Zusammenhang ist mit Verweis auf die Tagebuchstudienmethodik davon auszugehen, dass die Datenqualität mit zunehmender Retrospektivität u. a. aufgrund des zunehmenden zeitlichen Abstands zu einem erinnerten Ereignis abnimmt, was u. a. auf Erinnerungsverzerrungen zurückzuführen ist. Praktikabler sind hierbei schwach retrospektive Erhebungen, welche als ökologisch valider einzuschätzen sind als Erhebungen mit stark ausgeprägter Retrospektivität (vgl. Rausch, Kögler & Laireiter, 2012).

Die gefundenen deskriptiven Skalenwerte bewegen sich erwartungskonform eher im Skalenmittel (Abb. 2), was weder auf eine besonders schlechte, noch auf eine besonders gute Ausprägung der wahrgenommenen Bewertungstransparenz hinweist. Auffallend sind jedoch die durchgängig hohen Standardabweichungen, welche auf eine große Streuung der wahrgenommenen Transparenz zwischen den befragten Personen hindeuten. Diese ersten deskriptiven Ergebnisse lassen es lohnenswert erscheinen, das Thema der transparenten Leistungsbewertung im Referendariat in weiterführenden Studien intensiver zu behandeln. So stellen sich ausgehend von den gefundenen hohen Streuungen beispielsweise Anschlussfragen nach spezifischen Merkmalen besonders guter und besonders schlechter Seminarstandorte bzw. Ausbildungsschulen. Dabei wäre auch die Rolle von Seminarlehrkraft, Betreuungslehrkraft, Seminarvorstand und Schulleitung im Zusammenhang mit der wahrgenommenen Bewertungstransparenz zu untersuchen, ebenso wie der Zusammenhang mit strukturellen Merkmalen der Ausbildung sowie der Ausbildungsstandorte. In diesem Kontext stellt sich ebenfalls die Frage nach der Rolle personenspezifischer Eigenschaften der Referendare/-innen bei der Wahrnehmung von Bewertungstransparenz, wie z. B. die persönliche Belastbarkeit oder die Verfügbarkeit sozialer Ressourcen (vgl. Warwas, Neubauer & Panzer, 2016). Die weiterführende Untersuchung dieses Themenkomplexes erscheint v. a. auch deshalb von Relevanz, da sich die für die Lehrer/-innenbildung zuständigen staatlichen Einrichtungen der Bedeutung einer transparenten Bewertungspraxis durchaus bewusst zu sein scheinen. Strietholt und Terhart (2009) berichten hierzu über grundsätzlich verfügbare und damit einsehbare Beurteilungsinstrumente, welchen es jedoch u. a. an einer zufriedenstellenden testdiagnostischen Qualität mangelt. Mit Sicht auf unsere Ergebnisse drängt sich jedenfalls

die Vermutung auf, dass die seitens der staatlichen Einrichtungen verfügbar gemachten Beurteilungsbögen und Leistungsbeschreibungen zum gegenwärtigen Zeitpunkt noch nicht wie gewünscht zu einer durchgängigen und vor allem seitens der angehenden Lehrpersonen wahrgenommenen Transparenz beitragen. Bezüglich des gewünschten konstruktiven Feedbacks, ist anzumerken, dass dieses vorrangig durch die zuständige Seminarlehrkraft erfolgt, welche sich jedoch in einem ambivalenten Verhältnis zu den Referendaren/-innen befindet und aufgrund dessen der diesbezüglichen Forderung nach einem offenen, nachvollziehbaren und konstruktiven Feedback vermutlich nicht immer gerecht werden kann: Einerseits soll die Seminarlehrkraft die Rolle eines unterstützenden Coaches des/der auszubildenden Studienreferendars/-in einnehmen, andererseits fungiert sie vielerorts jedoch als eine beurteilende Instanz im Rahmen der Lehrprobe, des Kolloquiums, der mündlichen Prüfung sowie der schriftlichen Hausarbeit (vgl. Klusmeyer & Kehl, 2009; Warwas, Neubauer & Panzer, 2016).

Limitationen unserer Studie beziehen sich in erster Linie auf die Stichprobe. Zum einen konnten wir nur auf eine verhältnismäßig kleine Stichprobe zurückgreifen. Zum anderen besteht die Möglichkeit, dass es sich bei den untersuchten Personen dahingehend um eine selektive Stichprobe handelt, indem schwerpunktmäßig nur diejenigen an der Befragung teilgenommen haben, welche mit ihrer Ausbildungssituation unzufrieden sind bzw. waren. Dagegen sprechen jedoch die in Relation zu den jeweiligen Mittelwerten hohen Standardabweichungen bei den abgefragten Transparenzkriterien sowie bei der Zufriedenheit mit der Ausbildung. Weiterhin ist unsere Stichprobe nicht repräsentativ für die Lehramtsausbildung in Deutschland, da der Großteil der Teilnehmer/-innen das Referendariat in Bayern absolvierte und auch die verschiedenen Schularten nicht entsprechend der erwarteten Häufigkeiten vertreten sind (vgl. Statistisches Bundesamt, 2018). Entsprechende stichprobenbezogene Einschränkungen sind hierbei möglicherweise dem zugrundeliegenden Schneeballverfahren geschuldet, da sich das berufsbezogene Netzwerk von (angehenden) Lehrpersonen aufgrund der für die betreffenden Personen geltenden landeshoheitlichen (Ausbildungs-)Regelungen vermutlich weitgehend auf Personen beschränkt, die im gleichen Bundesland tätig sind. Eine bundesland- und schulartübergreifende Erhebung würde in diesem Zusammenhang bundesland- sowie schulartspezifische Vergleiche hinsichtlich der wahrgenommen Bewertungstransparenz ermöglichen. Bei einer ausreichend großen Stichprobe wären hierbei auch Analysen bzgl. schulspezifischer Unterschiede denkbar, inkl. der Identifikation schulspezifischer Faktoren, welche zu einer besonders guten bzw. schlechten Bewertungstransparenz im Referendariat beitragen. Eine weitere Limitation unserer Studie ist darin zu sehen, dass anhand der eingesetzten Instrumente keine Prüfung der Konstruktvalidität im eigentlichen Sinne möglich ist. Mittels der erhobenen Außenkriterien der Zufriedenheit mit der Ausbildung im Referendariat sowie der wahrgenommenen Belastung im Zusammenhang mit der Bewertungstransparenz lassen sich lediglich erste indirekte Hinweise durch entsprechende Korrelate finden, welche jedoch in weiterführenden Studien genauer zu prüfen wären.

Weitere Forschungsbemühungen in dem betrachteten Feld erscheinen jedoch relevant und lohnenswert. So ergeben sich für den von uns entwickelten Fragebogen

dergestalt Implikationen, dass dieser im Rahmen der Referendars/-innenausbildung zu Evaluationszwecken eingesetzt werden kann, um auf dessen Grundlage Faktoren zu identifizieren, welche in spezifischen Fällen zu einer besonders guten bzw. schlechten Bewertungstransparenz beitragen, um im letzteren Fall wiederum Ansatzpunkte für eine Optimierung der zweiten Phase der Lehramtsausbildung zu finden.

Literatur

Beiner, F. (1982). *Prüfungsdidaktik und Prüfungspsychologie. Leistungsmessung und Leistungsbewertung in der öffentlichen Verwaltung sowie in der beruflichen und allgemeinen Bildung.* Köln: Heymann.

Bohl, T. (2003). Neuer Unterricht – neue Leistungsbewertung. Grundlagen und Kontextbedingungen eines veränderten Bewertungsverständnisses. In O. Vorndran & D. Schnoor (Hrsg.), *Schulen für die Wissensgesellschaft. Ergebnisse des Netzwerkes Medienschulen* (S. 211–231). Gütersloh: Bertelsmann.

Bohl, T. (2004). Theoretische Strukturierung – Begründung neuer Beurteilungsformen. In H.-U. Grunder & T. Bohl (Hrsg.), *Neue Formen der Leistungsbeurteilung in den Sekundarstufen I und II* (S. 9–49). Baltmannsweiler: Schneider Hohengehren.

Brouër, B. (2014). *Selbstbeurteilung in selbstorganisationsoffenen Lernumgebungen. Entwicklung eines Modells der Selbstbeurteilung und evidenzbasierter Impulse für die Förderung der Selbstbeurteilung in der Praxis.* Bad Heilbrunn: Klinkhardt.

Dietrich, F. (2018). Ungewissheit im Referendariat. Professionalisierung als Krisenbearbeitung. In A. Paseka, M. Keller-Schneider & A. Combe (Hrsg.), *Ungewissheit als Herausforderung für pädagogisches Handeln* (S. 277–297). Wiesbaden: Springer VS.

Döbrich, P., & Abs, H. J. (2008). Evaluation der zweiten Phase der Lehrerbildung. Pädagogische Entwicklungsbilanzen mit Studienseminaren in Hessen. Frankfurt a. M.: DIPF. (zitiert in Strietholt & Terhart, 2009)

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften.* Berlin/Heidelberg: Springer

Drüge, M., Schleider, K., & Rosati, A.-S. (2014). Psychosoziale Belastungen im Referendariat – Merkmale, Ausprägungen, Folgen. *DDS – Die Deutsche Schule, 106*(4), 358–372.

Hartog, P., & Rhodes E. C. (1974). Prüfungszensuren in Geschichte und Englisch. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung – Texte und Untersuchungsberichte* (S. 78–89). Weinheim: Beltz.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81–112.

Heid, H. (1992). Was „leistet" das Leistungsprinzip? *Zeitschrift für Berufs- und Wirtschaftspädagogik, 88*(2), 91–108.

Heid, H. (2012). Der Beitrag des Leistungsprinzips zur Rechtfertigung sozialer Ungerechtigkeit. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete, 81*(1), 22–34.

Ingenkamp, K. (Hrsg.) (1974). *Die Fragwürdigkeit der Zensurengebung – Texte und Untersuchungsberichte.* Weinheim: Beltz.

Jürgens, E. (1998). *Leistung und Beurteilung in der Schule – Eine Einführung in Leistungs- und Bewertungsfragen aus pädagogischer Sicht.* Sankt Augustin: Academia.

Jürgens, E., & Sacher, W. (2000). *Leistungserziehung und Leistungsbeurteilung. Schulpädagogische Grundlegung und Anregungen für die Praxis.* Neuwied: Luchterhand.

Kaluza, G. (2015). *Stressbewältigung. Trainingsmanual zur psychologischen Gesundheitsförderung.* Berlin/Heidelberg: Springer.

Klafki, W. (1996). Sinn und Unsinn des Leistungsprinzips in der Schule. In W. Klafki (Hrsg.), *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik* (S. 209–247). Weinheim: Beltz.

Klein, H. (2009). *Transparente Leistungsbewertung und Zensurengebung in der Sekundarstufe I. Umfassend – praxisorientiert – fair.* Baltmannsweiler: Schneider Hohengehren.

Klusmann, U., Kunter, M., Voss, T., & Baumert, J. (2012). Berufliche Beanspruchung angehender Lehrkräfte: Die Effekte von Persönlichkeit, pädagogischer Vorerfahrung und professioneller Kompetenz. *Zeitschrift für Pädagogische Psychologie, 26*(4), 275–290.

Klusmeyer, J., & Kehl, V. (2009). Zur Professionalität der Ausbildungslehrer an berufsbildenden Schulen. *bwp@ Berufs- und Wirtschaftspädagogik – online.* http://www.bwpat.de/profil2/klusmeyer_kehl_profil2.shtml [25. 07. 2018].

KMK = Kultusministerkonferenz (2012). *Ländergemeinsame Anforderungen für die Ausgestaltung des Vorbereitungsdienstes und die abschließende Staatsprüfung.* http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2012/2012_12_06-Vorbereitungsdienst.pdf [20. 12. 2017].

KMK = Kultusministerkonferenz (2013). *Übersicht zu den Lehramtsprüfungen und -befähigungen in den Ländern der Bundesrepublik Deutschland.* https://www.kmk.org/dokumentation-statistik/rechtsvorschriften-lehrplaene/uebersicht-lehrerpruefungen.html [20. 12. 2017].

KMK = Kultusministerkonferenz (2014). *Standards für die Lehrerbildung: Bildungswissenschaften.* Beschluss der Kultusministerkonferenz vom 16. 12. 2004 i. d. F. vom 12. 06. 2014. http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf [06. 02. 2019].

Košinár, J. (2014). Typenspezifischer Umgang mit den Anforderungen des Referendariats. *Lehrerbildung auf dem Prüfstand, 7*(2) 120–137.

Košinár, J. (2018). Konstruktionen von Professionalität und Ungewissheitserfahrungen im Referendariat. In A. Paseka, M. Keller-Schneider, & A. Combe (Hrsg.), *Ungewissheit als Herausforderung für pädagogisches Handeln* (S. 255–275). Wiesbaden: Springer VS.

LLPA = Landeslehrerprüfungsamt in Baden-Württemberg (2018). *Verordnungen über die Vorbereitungsdienste und Zweiten Staatsprüfungen für die verschiedenen Lehrämter,* http://www.llpa-bw.de/,Lde/Startseite/Pruefungsordnungen/Zweite-Staatspruefungen [06. 02. 2019].

LPO II = Lehramtsprüfungsordnung II (2004). *Ordnung der Zweiten Staatsprüfung für ein Lehramt an öffentlichen Schulen vom 28. Oktober 2004 (GVBl. S. 428), BayRS 2038-3-4-8-11-K,* http://www.gesetze-bayern.de/Content/Document/BayLPO_II [06. 02. 2019].

Meyer, H. (2004). *Was ist guter Unterricht?* Berlin: Cornelsen.

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus User's Guide.* Sixth Edition. Muthén & Muthén, Los Angeles, CA. https://www.statmodel.com/download/usersguide/Mplus%20Users%20Guide%20v6.pdf [06. 02. 2019].

Oehmig, B. (2005). *Erfolgreiches Referendariat. Hinweise und Hilfen für Schule und Lehramtspraktika.* Frankfurt a. M.: Lang.

Oser, F., Hascher, T., & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie des „negativen" Wissens. In A. Wolfgang (Hrsg.), *Fehlerwelten: Vom Fehlermachen und Lernen aus Fehlern. Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlaß des 60. Geburtstags von Fritz Oser* (S. 11–41). Opladen: Leske + Budrich.

Paradies, L., Wester, F., & Greving, J. (2005). *Leistungsmessung und -bewertung.* Berlin: Cornelsen.

Rausch, A., Kögler, K., & Laireiter, A.-R. (2012). Editorial. In A. Rausch, K. Kögler, & A.-R. Laireiter (Hrsg.), *Tagebuchverfahren zur prozessnahen Datenerhebung in Feldstudien – Gestaltungsparameter und Anwendungsempfehlungen.* Themenheft *Empirische Pädagogik, 26*(2), 183–199.

Rheinberg, F. (2001). Bezugsnormen und schulische Leistungsbeurteilung. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 59–71). Weinheim: Beltz.

Rost, D.H., & Schermer, F.J. (2010). Leistungsängstlichkeit. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 405–413). Weinheim: Beltz.

Rudow, B. (2011). *Die gesunde Arbeit.* München: Oldenbourg Wissenschaftsverlag.

Saldern, M.v. (1999). *Schulleistung in Diskussion.* Baltmannsweiler: Schneider Hohengehren.

Schaarschmidt, U. (2011). Potsdamer Lehrerstudie – ein erstes Fazit. In U. Schaarschmidt (Hrsg.), *Halbtagsjobber? Psychische Gesundheit im Lehrerberuf – Analyse eines veränderungsbedürftigen Zustandes?* (S. 141–156). Weinheim: Beltz.

Scheuch, K., Haufe, E., & Seibt, R. (2015). Lehrergesundheit. *Deutsches Ärzteblatt, 112*(20), 347–356.

Schrader, F.-W., & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45–58). Weinheim: Beltz.

Starch D., & Elliot E.C. (1974). Die Verlässlichkeit der Zensuren von Mathematikarbeiten. In K. Ingenkamp (Hrsg.), *Die Fragwürdigkeit der Zensurengebung – Texte und Untersuchungsberichte* (S. 69–77). Weinheim: Beltz.

Statistisches Bundesamt (2018). Lehrkräfte nach Schularten und Beschäftigungsumfang, Schuljahr 2016/17. https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Schulen/Tabellen/AllgemeinBildendeBeruflicheSchulenLehrkraefte.html;jsessionid=963E57BEBAFFA4E40DC9BB20EBDD8C11.InternetLive1 [06.02.2019].

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.

Stritzke, R., Harles, M., & Wenzel, K. (1982). *Schüler kennen und beurteilen – Arbeitshilfen zur Erstellung pädagogischer Wortgutachten und von Gutachten zur Leistungsbeurteilung.* Donauwörth: Ludwig Auer.

Warwas, J., Neubauer, J., & Panzer, E. (2016). Unterstützung und Beanspruchung im Referendariat aus der Perspektive angehender Berufsschullehrkräfte. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 112*(2), 294–313.

Wengert, H.G. (2006). Leistungsbeurteilung in der Schule. In G. Bovet & V. Huwendiek (Hrsg.), *Leitfaden Schulpraxis. Pädagogik und Psychologie für den Lehrberuf* (S. 293–319). Berlin: Cornelsen.

Winne, P., & Butler, D. (1994). Student Cognition in Learning from Teaching. In T. Husén & T. Postlewaite (Eds.), *The International Encyclopedia of Education* (pp. 5738–5745). Oxford: Pergamon.

Abstract: Existing research on assessment instruments within teacher training reveals that assessment practices are not always transparent (e.g., Strietholt & Terhart, 2009). Moreover, there is little known about how transparent teachers in training perceive their training to be. Thus, we present a questionnaire that can be used to measure the perceived assessment transparency within teacher training. First, we identified (1) understandable and clearly communicated achievement expectations and assessment criteria, and (2) constructive feedback as the two core dimensions of perceived assessment transparency. Second, we developed questionnaire items that are related to both dimensions and that are additionally matched to five assessment situations within teacher training (e.g., demonstration lessons, oral examinations). The results of factor analyses support the assumption of two subscales ("transparent achievement expectations and assessment criteria", "constructive feedback") for each of the considered assessment situations within teacher training. The found scale values are in the middle range. Neither the perceived assessment transparency is very good or very poor in the sample. However, the high standard deviations are remarkable and indicate high variation within the sample. As a result of our study, our questionnaire can be used in teacher training for evaluation purposes in order to identify potential opportunities for the optimisation of teacher training.

Keywords: Achievement Evaluation, Teacher Training, Assessment Criteria, Feedback, Transparency

Anschrift der Autoren_innen

Dr. rer. pol. Dr. habil. Tobias Kärner, Universität Konstanz,
Juniorprofessur für Wirtschaftspädagogik mit dem Schwerpunkt Heterogenität,
Fach 123, 78457 Konstanz, Deutschland
E-Mail: tobias.kaerner@uni-konstanz.de

Dr. phil., Dipl.-Päd. Caroline Bonnes, Universität Konstanz,
Wissenschaftliche Mitarbeiterin am Lehrstuhl für Betriebspädagogik,
Fach 125, 78457 Konstanz, Deutschland
E-Mail: caroline.bonnes@uni-konstanz.de

M.Sc. Christian Schölzel, Studienrat an der Städtischen Wirtschafts- und Berufsoberschule
Schwabach, Südliche Ringstraße 9a, 91126 Schwabach, Deutschland
E-Mail: c.schoelzel@ws-schwabach.de