

JAHRBUCH FÜR  
**ALLGEMEINE DIDAKTIK**

**2017**

**Thementeil**  
Allgemeine Didaktik und  
Lehrer/innenbildung





# JAHRBUCH FÜR ALLGEMEINE DIDAKTIK

2017

## Editorial Board

Karl-Heinz Arnold (Hildesheim), Thorsten Bohl (Tübingen),  
Ilona Esslinger-Hinz (Heidelberg), Michaela Gläser-Zikuda (Erlangen-Nürnberg),  
Ludwig Haag (Bayreuth), Ulrike Hanke (Freiburg), Stefan T. Hopmann (Wien),  
Brian Hudson (Sussex), Manuela Keller-Schneider (Zürich), Ewald Kiel  
(München), Barbara Koch-Priewe (Bielefeld), Frank Lipowsky (Kassel), Gabi  
Reinmann (Hamburg), Kurt Reusser (Zürich), Norbert M. Seel (Freiburg),  
Matthias Trautmann (Siegen), Klaus Zierer (Augsburg)

## Zusammenstellung und Betreuung des Thementails

Manuela Keller-Schneider, Michaela Gläser-Zikuda und Matthias Trautmann

## Geschäftsführender Herausgeber

Prof. Dr. Klaus Zierer  
Dorthe Behrens (Assistenz)  
Universität Augsburg  
Universitätsstraße 10  
86159 Augsburg

## Homepage

[www.paedagogik.de](http://www.paedagogik.de)



Gedruckt auf umweltfreundlichem Papier (chlor- und säurefrei hergestellt).

**Bibliografische Information der Deutschen Nationalbibliothek**

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

ISBN: 978-3-8340-1822-9

Schneider Verlag Hohengehren, 73666 Baltmannsweiler

Homepage: [www.paedagogik.de](http://www.paedagogik.de)

Alle Rechte, insbesondere das Recht der Vervielfältigung sowie der Übersetzung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung des Verlages reproduziert werden.

© Schneider Verlag Hohengehren, Baltmannsweiler 2017.

Printed in Germany. Druck: Esser, Bretten

## Thementeil

### Allgemeine Didaktik und Lehrer/innenbildung

MANUELA KELLER-SCHNEIDER / MICHAELA GLÄSER-ZIKUDA / MATTHIAS TRAUTMANN <b>Einführung in den Thementeil ‚Allgemeine Didaktik und Lehrer/innenbildung‘</b> .....	S. 7
MIRIAM HESS / KATHARINA WERKER / FRANK LIPOWSKY <b>Was wissen Lehramtsstudierende über gutes Feedback?</b> .....	S. 11
SILVIA GREITEN <b>Unterrichtsplanung im Praxissemester – zwischen Intuition, Phasenrastern und Wissen?</b> .....	S. 30
ELFRIEDE AMTMANN / JOHANNES BRANDAU / URSULA GRASSER / NORBERT HOLZER <b>Bewältigungsstrategien bei einem ethisch-pädagogischen Dilemma in der Kompetenzentwicklung angehender Lehrkräfte</b> .....	S. 47
ANNA PINEKER-FISCHER <b>Aufgabenorientiertes Peer-Tutorium zur Professionalisierung im Modul Deutsch als Zweitsprache</b> .....	S. 68
KERSTIN GÖBEL / KATHARINA NEUBER <b>Potenziale der Einholung von Schülerrückmeldungen für die Unterrichtsreflexion in den Phasen des Lehrerberufs</b> .....	S. 88
SEBASTIAN JÜNGER / CHRISTIAN REINTJES <b>Lehrer/innenbildung im hybriden Raum</b> .....	S. 102
HENRIKE KOPMANN / HORST ZEINZ <b>Lehrer/innenbildung für eine inklusive Schule</b> .....	S. 122

MANUELA KELLER-SCHNEIDER / ELIF ARSLAN / JASPER MAAS / UWE HERICKS <b>Die Wahrnehmung von Anforderungen zur Planung von Unterricht und ihre Bezüge zu den beruflichen Entwicklungsaufgaben</b> .....	S. 138
SUSANNE PREDIGER / TIMO LEUDERS / BETTINA RÖSKEN-WINTER <b>Drei-Tetraeder-Modell der gegenstandsbezogenen Professionalisierungsforschung</b> .....	S. 159
MICHAELA GLÄSER-ZIKUDA / FLORIAN HOFMANN <b>Rezension</b> .....	S. 178
<b>Impressum</b> .....	S. 184

**Thementeil**  
Allgemeine Didaktik und  
Lehrer/innenbildung



MANUELA KELLER-SCHNEIDER / MICHAELA GLÄSER-ZIKUDA /  
MATTHIAS TRAUTMANN

## Einführung in den Thementeil ‚Allgemeine Didaktik und Lehrer/innenbildung‘

### Editorial

In Zeiten einer sich verändernden Schul- und Hochschullandschaft ist auch die Lehrerinnen- und Lehrerbildung erneut zum Gegenstand zahlreicher Reformvorschläge und Reforminitiativen geworden.

In *Deutschland* wird der Wandel beispielsweise in einem Ausbau von Praxisanteilen im Rahmen der ersten Phase der Lehrer/innenbildung manifest. Im Gegensatz dazu – im Kontext der Gründung der Pädagogischen Hochschulen seit der Jahrtausendwende – ist dieser Wandel in der *Schweiz* und in *Österreich* durch eine stärkere Wissenschaftsorientierung der Lehrer/innenbildung charakterisiert. Beiden Bestrebungen gemeinsam ist eine Verknüpfung von allgemeinem, theoriegestütztem Wissen mit spezifischen, erfahrungsbezogenen Kenntnissen, um Erfahrungen theoriegestützt und Theorien erfahrungsbezogen zu diskutieren sowie die Handlungs- und Reflexionskompetenz der angehenden Lehrpersonen zu fördern. Lehrer/innenbildung – ob ein- oder zweiphasig strukturiert – soll notwendige und geeignete Grundlagen für die Professionalisierung legen.

Lehrpersonen sind ein Berufsleben lang gefordert, sich eigenaktiv und selbstreguliert mit sich wandelnden Aufgaben und Anforderungen auseinanderzusetzen. Die Professionalisierung von Lehrpersonen gewinnt damit über die Grundausbildung hinaus an Bedeutung. Die Frage nach lernwirksamen didaktischen Arrangements ist somit nicht nur für die Grundausbildung, sondern auch für die Fort- bzw. Weiterbildung zentral. Zu klären ist, welche institutionellen Strukturen, Lernarrangements und Interaktionen zwischen unterschiedlichen Akteur/innen zur Professionalisierung von Lehrpersonen beitragen.

Das vorliegende Heft greift diese Thematik auf und umfasst Beiträge, die sich theoretisch und/oder empirisch mit allgemein- und fachdidaktischen Fragen bezüglich der Veränderungen in der Lehrer/innenbildung auf *individueller*, *interaktioneller* sowie *institutioneller* Ebene auseinandersetzen. Dabei wird der Begriff einer ‚Allgemeinen Didaktik‘ über die Unterrichtseinheit hinaus auf die Strukturierung von Prozessen auf unterschiedlichen Ebenen der Schule sowie auf mehrere Phasen der Lehrer/innenbildung erweitert.

In den Beiträgen des ersten Teils stehen *Professionalisierungsprozesse von Studierenden als angehende Lehrpersonen* im Zentrum. Der Beitrag von *Miriam Hess, Katharina Werker und Frank Lipowsky* mit dem Titel **„Was wissen Lehramtsstudierende über gutes Feedback?“** fokussiert auf die Frage des Wissens, welches Grundlagen für die Entwicklung von Handlungskompetenz bietet. Am Erwerb von konzeptionellem Wissen über ein gutes Feedback zur Förderung der Schüler/innen bezogen auf ihre Lernprozesse wird erläutert, inwiefern eine Lehrveranstaltung zum Wissenserwerb beitragen kann. Die Ergebnisse der Studie zeigen, dass der Lernzuwachs der Studierenden über einen Wissenstest nachgewiesen werden kann und dass die Ergebnisse des Wissenstests jedoch nur sehr schwach mit dem Umfang des selbsteingeschätzten Wissens korrelieren.

Der Beitrag von *Silvia Greiten* mit dem Titel **„Unterrichtsplanung im Praxissemester – zwischen Intuition, Phasenrastern und Wissen?“** untersucht, wie Lehramtsstudierende im Praxissemester in der Planung von Unterricht vorgehen und inwiefern sie auf erworbenes Wissen zurückgreifen. Im Zentrum des Beitrags steht somit die Verbindung zwischen Theorie und Praxis am Beispiel der Planung von Unterricht. Die Ergebnisse ergeben, dass die Planung von Unterricht den Studierenden wichtig ist und sie diese auch als Messlatte der Qualität der eigenen Arbeit betrachten, dass sie sich in der Planung des Unterrichts jedoch wenig auf erworbenes Wissen abstützen.

Im Beitrag **„Bewältigungsstrategien bei einem ethisch-pädagogischen Dilemma in der Kompetenzentwicklung angehender Lehrkräfte“** von *Elfriede Amtmann, Johannes Brandau, Ursula Grasser und Norbert Holzer* wird der These nachgegangen, dass ethisches Urteilsvermögen als fächerübergreifende Kompetenz einen wesentlichen Bestandteil einer zukunftsorientierten Lehrer/innenbildung darstellt. Über Fall vignetten identifizierte Befunde zur Entwicklung dieser Kompetenz zeigen, dass sich Studierende des 6. Semesters bezüglich ihrer ethischen Urteilskompetenz nicht von Studierenden des 2. Semesters unterscheiden, dass sich jedoch individuell differente Entwicklungen erkennen lassen, die unabhängig von der Entwicklung pädagogisch-psychologischer und didaktischer Kompetenzen verlaufen. Dabei stellt sich die Frage, inwiefern ethische Urteilskompetenz lernbar ist oder welche Settings hierfür erforderlich wären.

In den folgenden drei Beiträgen wird das *Mitwirken weiterer Akteure in der Professionalisierung angehender Lehrpersonen* untersucht, wobei die Beiträge von Schüler/innen, Tutor/innen und Mentor/innen thematisiert werden.

*Anna Pineker-Fischer* stellt in ihrem Beitrag mit dem Titel **„Aufgabenorientiertes Peer-Tutorium zur Professionalisierung im Modul Deutsch als Zweitsprache“** dar, wie Professionalisierung im Grundlagenmodul Deutsch als Zweitsprache an der Universität Duisburg-Essen realisiert wird. Empirisch wird untersucht, welche Wirkung ein Peer-Tutorium auf beliefs von Lehramtsstudierenden hat, wie sich das Feedback der Tutor/innen gestaltet und als wie hilfreich es bewertet wird.

*Kerstin Göbel und Katharina Neuber* zeigen in ihrem Beitrag **„Potenziale der Einholung von Schülerrückmeldungen für die Unterrichtsreflexion in den Phasen des Lehrerberufs“**, dass Lehramtsstudierende sowohl zu Beginn als auch am Ende des Praxissemesters Rückmeldungen von Schüler/innen als nützlich erachten sowie Bereitschaft zeigen, diese zu nutzen, wobei die individuelle Umsetzung des Einholens von Rückmeldungen bei den Schüler/innen während des Praxissemesters abnimmt.

Im Beitrag **„Lehrer/innenbildung im hybriden Raum“** von *Sebastian Jünger und Christian Reintjes* wird die Kooperation zwischen Studierenden und ihren Auszubildenden von Hochschule und Schulfeld beleuchtet. Dem Konzept des Relationierens folgend werden Erfahrungen, Tätigkeiten, Kommunikationen und Produkte an den Lernorten Schule und Hochschule zueinander in Bezug gesetzt, wobei den Mentorierenden sowohl als Lehrenden als auch als Lernenden eine wichtige Rolle zugeschrieben wird.

Die dritte Gruppe von Beiträgen fokussiert auf die Institution als Lernort, die mit ihren Strukturen und Praktiken zur Professionalisierung von Lehrpersonen beiträgt.

*Henrike Kopmann und Horst Zeinz* thematisieren in ihrem Beitrag „**Lehrer/innenbildung für eine inklusive Schule**“ die Frage nach der Bedeutung von Strukturen der ersten, universitären Ausbildungsphase sowie der zu erwerbenden Kompetenzen von Lehrpersonen im Kontext inklusiver Bildungsprozesse. Ergebnisse einer Befragung heben insbesondere die Fähigkeiten eines adaptiven, auf die individuellen Schüler/innen eingehenden Vorgehens hervor, welche didaktische Kompetenzen sowie Kompetenzen bezüglich Klassenführung und Lernbegleitung erfordern – und damit Kompetenzen umschreiben, die für alle Lehrpersonen von Bedeutung sind.

Der Beitrag von *Manuela Keller-Schneider, Elif Arslan, Jasper Maas und Uwe Hericks* mit dem Titel „**Die Wahrnehmung von Anforderungen zur Planung von Unterricht und ihre Bezüge zu den beruflichen Entwicklungsaufgaben**“ thematisiert über einen Vergleich zwischen deutschen und schweizerischen Lehrpersonen die Frage der Bedeutung von institutionellen Ausbildungskontexten mit je spezifischen allgemeindidaktischen Schwerpunktsetzungen für die Professionalisierung von angehenden Lehrpersonen. Die Befunde der Studie zeigen, dass das latente Verständnis von Berufsanforderungen bei Ausbildungsabschluss zwischen Abgänger/innen einer ein- bzw. zweiphasigen Ausbildung differiert und dass sich die identifizierten Spezifika im Verlauf der ersten zwei Berufsjahre angleichen.

Der Thementeil schließt mit dem Beitrag von *Susanne Prediger, Timo Leuders und Bettina Rösken-Winter* „**Drei-Tetraeder-Modell der gegenstandsbezogenen Professionalisierungsforschung**“. Darin wird ein spezifisches didaktisches Modell vorgestellt, welches auf drei Ebenen ansetzt und das unterrichtliche Geschehen sowohl auf der Ebene des Unterrichts mit den Schüler/innen fokussiert als auch auf der Ebene von Fortbildung und weiterführender Qualifikation. Die Lehrperson der einen Ebene ist dabei zugleich auch Lernende auf der je übergeordneten Ebene und nimmt dabei eine wechselnde Position im didaktischen Dreieck ein. Dieses wird um den Zugang über Materialien und Medien zum Tetraeder erweitert und soll Leitlinien zur Beforschung des Geschehens auf allen drei Ebenen bieten.

Mit dem vorliegenden Themenheft wird der Versuch unternommen, die aktuellen Veränderungen in der Lehrer/innenbildung anhand ausgewählter theoretisch-konzeptioneller Beiträge sowie empirischer Studien zur Diskussion zu stellen. Insgesamt wurden fünfzehn Beiträge eingereicht, von denen neun nach einem peer-review-Verfahren positiv begutachtet wurden. Damit diese Beiträge in der nun veröffentlichten finalen Version publiziert werden konnten, haben zahlreiche Kolleginnen und Kollegen als Reviewer/innen mitgewirkt: Martin Drahm (Tübingen), Alexander Gröschner (Jena), Ewald Kiel (München), Katja Kansteiner (Weingarten), Marc Kleinknecht (Lüneburg), Barbara Koch-Priewe (Bielefeld), Kathrin Krammer (Luzern), Annelies Kreis (Zürich), Wolfgang Meseth (Marburg), Carsten Rohlf (Heidelberg), Niclas Schaper (Paderborn), Stefanie Schnebel (Weingarten), Ulrike Stadler-Altmann (Bozen), Kathrin Uplegger (Berlin), Klaus Zierer (Augsburg).

Ihnen allen sei ein großer Dank ausgesprochen.

Unser Dank gilt auch allen Autorinnen und Autoren, die an der siebten Ausgabe des Jahrbuchs für Allgemeine Didaktik mitgewirkt haben, Dorthe Behrens (Oldenburg) für die Aufbereitung der Beiträge, sowie dem Verlag, namentlich Rainer Schneider, für die Unterstützung.

Zürich, Erlangen-Nürnberg und Siegen, im Oktober 2017

Manuela Keller-Schneider, Michaela Gläser-Zikuda und Matthias Trautmann

Herausgeber/innen des Themenhefts JfAD 2017

MIRIAM HESS / KATHARINA WERKER / FRANK LIPOWSKY

## Was wissen Lehramtsstudierende über gutes Feedback?

### Zur Erfassung konzeptuellen Wissens und zu dessen Zusammenhang mit der Selbsteinschätzung der Studierenden

#### Abstract

Der hohen Bedeutsamkeit von Feedback für den Lernprozess von Schülerinnen und Schülern stehen Ergebnisse aus empirischen Studien gegenüber, die zeigen, dass Lehrpersonen selten lernförderliches Feedback geben. Es ist davon auszugehen, dass die Fähigkeit von Lehrpersonen, förderliches Feedback zu erteilen, grundlegend von deren konzeptuellem Wissen über Merkmale guten Feedbacks abhängt. Daher wurde in universitären Seminaren versucht, das Wissen von Lehramtsstudierenden zum Thema Feedback zu fördern. Zur Erfassung des konzeptuellen Wissens wurde ein Instrument entwickelt, dessen ausführliche Vorstellung im Mittelpunkt des vorliegenden Beitrags steht. Es wird gezeigt, dass das Wissen über Feedback mit dem Instrument reliabel und objektiv erfasst sowie detailliert beschrieben werden kann.  $N = 54$  Lehramtsstudierende besuchten Seminare zum Thema Feedback und bearbeiteten vorher und nachher sowohl den Test als auch Fragen zur Selbsteinschätzung ihres Wissens(zuwachses). Es zeigt sich, dass das mit dem Test erhobene Wissen und die generell sehr positiv ausgeprägte Selbsteinschätzung der Studierenden kaum miteinander zusammenhängen. Dies unterstreicht die Bedeutung einer testbasierten Erfassung konzeptuellen Wissens.

#### Keywords

Lehrerinnen- und Lehrerwissen, Hochschullehre, Lehrerinnen- und Lehrerbildung, Feedback, Selbsteinschätzung

## Einleitung

Obwohl Feedback mit einer Effektstärke von  $d = 0.73$  eines der lernwirksamsten Unterrichtsmerkmale darstellt (Hattie, 2013), zeigen Studien, dass Lehrpersonen den Lernenden nur selten informative Rückmeldungen geben (z.B. Hattie & Timperley, 2007; Lotz, 2015). Daher sollte dem Thema möglichst bereits in der universitären Lehrerinnen- und Lehrerbildung besondere Beachtung geschenkt werden. Eine grundlegende Voraussetzung, um gutes Feedback zu geben, ist das Wissen, welche Kriterien gutes Feedback kennzeichnen.

Im Zuge des verstärkten Interesses an der Untersuchung und Steigerung der Qualität von Maßnahmen einer evidenzbasierten Lehrerinnen- und Lehrerbildung (z.B. Kunter & Pohlmann, 2015) ist die Entwicklung von Instrumenten zur Untersuchung des Lernzuwachses von Studierenden besonders bedeutsam. Neben der klassischen Lehrveranstaltungsevaluation sowie Befragungen der Studierenden zur Selbsteinschätzung sollten Methoden entwickelt werden, die es erlauben, den Lernzuwachs allgemein sowie themenspezifisch auf verschiedenen Ebenen systematisch zu überprüfen. Aufschnaiter und Blömeke stellten im Jahr 2010 hierzu fest, dass sich „die empirische Erfassung des mit Lehrerprofessionalität verbundenen Kompetenzgefüges [...] oft noch in den Anfängen“ (S. 365) befindet. Inzwischen liegen verschiedene Instrumente vor, die sich meist global auf das pädagogisch-psychologische Wissen von Studierenden beziehen, nicht aber auf inhaltlich konkrete Teilbereiche, die im Rahmen einzelner Lehrveranstaltungen erarbeitet werden können.

Das Ziel des vorliegenden Beitrags besteht daher darin, ein selbst entwickeltes Instrument zur Erfassung des konzeptuellen Wissens über Feedback vorzustellen. Berichtet wird anschließend, wie sich das konzeptuelle Wissen mit Hilfe des Instruments beschreiben lässt. Außerdem wird die Frage fokussiert, wie das durch den Test erfasste Wissen mit der Selbsteinschätzung des Wissens durch die Studierenden zusammenhängt. Die Daten stammen aus insgesamt drei Seminaren zum Thema Feedback, in deren Rahmen insgesamt 54 Lehramtsstudierende den Wissenstest sowie Fragen zur Selbsteinschätzung ihres Wissens jeweils vor und nach dem Besuch des Seminars bearbeiteten.

Bevor auf die Studie selbst eingegangen wird, wird die Bedeutung konzeptuellen Wissens für Lehrpersonen herausgearbeitet sowie die adäquate Erfassung professioneller Kompetenzen von Lehrpersonen diskutiert. Anschließend wird dargestellt, welche Merkmale nach dem aktuellen Forschungsstand gutes Feedback kennzeichnen, bevor der Forschungsstand zu studentischen Selbsteinschätzungen skizziert wird.

## **Konzeptuelles Wissen als Aspekt der professionellen Kompetenz von Lehrpersonen**

Als zentrale Aspekte professioneller Kompetenz von Lehrpersonen werden in der aktuellen Diskussion meist die Aspekte Wissen, Überzeugungen, Motivation und selbstregulative Fähigkeiten unterschieden (Kunter et al., 2011), wobei das Wissen neben den Überzeugungen als besonders relevant für das berufliche Handeln angesehen wird (Kunter & Pohlmann, 2015).

Dabei werden – neben metakognitivem Wissen – v.a. deklaratives und prozedurales Wissen unterschieden. Deklaratives Wissen („Wissen, dass“) kann sowohl Fakten als auch komplexeres Zusammenhangswissen umfassen. Dahingegen beschreibt prozedurales Wissen („Wissen, wie“) das Können, das nicht unbedingt direkt verbalisierbar ist. De Jong und Ferguson-Hessler (1996) grenzen von prozeduralem Wissen des Weiteren strategisches, situationales und konzeptuelles Wissen ab. Unter konzeptuellem Wissen verstehen sie statisches Wissen über Fakten, Begriffe und Prinzipien in einer bestimmten Domäne (vgl. auch Gruber & Stamouli 2015).

Die bekannteste Unterscheidung unterschiedlicher Wissensarten als Facetten professioneller Kompetenz von Lehrpersonen geht auf Shulman (1987) zurück, der „Content/Subject Matter Knowledge (CK)“, „Pedagogical Content Knowledge (PCK)“ und „Pedagogical Knowledge (PK)“ unterscheidet – also Fachwissen über den zu unterrichtenden Gegenstand, fachdidaktisches Wissen über die Instruktion fachlicher Inhalte sowie pädagogisches Wissen „über die Schaffung und Optimierung von Lehr-Lern-Situationen sowie entwicklungspsychologisches und pädagogisch-psychologisches Grundwissen“ (Kunter & Pohlmann, 2015, S. 265; vgl. auch Baumert & Kunter, 2006; König & Blömeke, 2012).

„Viele pädagogische Wissensbestände manifestieren sich erst in fachlichen Zusammenhängen“ (Aufschnaiter & Blömeke, 2010). Daher treten insbesondere zwischen fachdidaktischem und pädagogisch-psychologischem Wissen Überschneidungen auf (Aufschnaiter & Blömeke, 2010), die sich auch im Bereich Feedback zeigen. So lässt sich Wissen über Feedback einerseits dem allgemein-pädagogischen Wissen zuordnen, das für Lernprozesse in allen Fächern relevant ist. Um gutes Feedback zu geben, ist andererseits auch fachdidaktisches Wissen über Schülervorstellungen und typische Schülerfehler sowie Wissen

über fachspezifische Instruktionsstrategien relevant, was dem fachdidaktischen Wissen zugeordnet wird (Schmelzing, Wüsten, Sandmann & Neuhaus, 2008).

Der Wissenserwerb gilt als besonders erfolgreich, wenn die Wissensbestände hoch vernetzt sind und Anwendungsqualität besitzen. In diesem Fall wird häufig auch der Begriff der Kompetenz genutzt (Kunter & Pohlmann, 2015; Renkl, 2015). Terharts (2002) Standards für Lehrerbildung greifen diese Differenzierung auf und unterscheiden in einer Taxonomie die Kompetenzfacetten *Wissen, Reflektieren, Kommunizieren, Beurteilen* und *Können*. Das Können wird dabei vorwiegend als Ziel der zweiten Phase der Lehramtsausbildung gesehen (Baumert & Kunter, 2006).

Dass Wissen für das Können eine wichtige Voraussetzung darstellt, konnte in mehreren Studien belegt werden. Voss, Kunter, Seiz, Hoehne und Baumert (2014) konnten beispielsweise einen positiven Zusammenhang zwischen der von den Lernenden eingeschätzten Effizienz der Klassenführung und der konstruktiven Lernunterstützung angehender Lehrpersonen mit deren pädagogisch-psychologischem Wissen nachweisen. Zu ähnlichen Befunden gelangen auch König und Pflanzl (2016) sowie Lenke und Kollegen (2016), wobei letztere zusätzlich Effekte des Wissens auf die Schülerleistungen – vermittelt über die Unterrichtsqualität – zeigen können. Dabei ist anzunehmen, „dass das deklarative Wissen nicht unmittelbar handlungsleitend ist, sondern indirekt über den Prozess der Prozeduralisierung erst in handlungsleitendes Wissen übergeht“ (Lenke et al., 2016, S. 214).

## Diskussion um die adäquate Erfassung professioneller Kompetenzen von Lehrpersonen

Teilweise wurden in der Vergangenheit eher distale Kriterien zur Abschätzung des Wissens von Lehrpersonen herangezogen – wie die Dauer der Ausbildung oder die Anzahl besuchter Kurse (Kunter & Pohlmann, 2015). Im Zuge der neueren kompetenzorientierten Lehrerinnen- und Lehrerbildungsforschung wird in vielen Studien versucht, Aspekte professioneller Kompetenz zu modellieren und empirisch zu erfassen (Aufschnaiter & Blömeke, 2010). Das Ziel besteht dabei darin, „geeignete Maße zur Erfassung des Wissens zu entwickeln, die genutzt werden können, um Unterschiede zwischen Lehrkräften zuverlässig zu beschreiben“ (Kunter & Pohlmann, 2015, S. 265).

Insbesondere groß angelegte Studien – wie COACTIV, MT21 und TEDS-M (z.B. Blömeke, Kaiser & Lehmann, 2008, 2010a, b) haben sich mit der empirischen Erfassung professioneller Kompetenzen beschäftigt, wobei allerdings meist nicht pädagogisch-psychologisches Wissen über spezifische Themen untersucht wurde. Die entsprechenden Tests erfassen hingegen den Wissenszuwachs durch das Studium im Gesamten und sind dadurch breit angelegt (z.B. König & Seifert, 2012). Um den Wissenszuwachs in ausgewählten Inhaltsbereichen detaillierter prüfen zu können, sind diese Tests daher meist zu allgemein:

Mit zunehmendem Abstand vom Unterricht und seinem Kontext dürfte es schwieriger werden, die professionelle Handlungsrelevanz von allgemeinem pädagogischen Wissen auch nachzuweisen. Dies gilt vor allem für die bildungswissenschaftlichen Grundlagen, bei denen man nur von indirekt vermittelten Wirkungszusammenhängen ausgehen kann. (Baumert & Kunter, 2006, S. 484)

An Instrumente zur Erfassung pädagogisch-psychologischen Wissens stellten Voss und Kunter (2011, S. 202) für die COACTIV-Studie folgende Anforderungen: „(a) Das erfasste Wissen sollte nicht spezifisch für ein bestimmtes Fach, sondern für das Unterrichten in

verschiedenen Fächern relevant sein. (b) Das erfasste Wissen sollte direkte theoretisch angenommene Bedeutsamkeit für das Unterrichten haben. (c) Die korrekten Antworten der Multiple-Choice Fragen sowie die richtigen Antwortkategorien der offenen Fragen mussten [im Rahmen der COACTIV-Studie] normativ als richtig beurteilt werden können, wobei sich der normative Gehalt aus empirischen Forschungsbefunden speiste.“

Zur Erfassung des Professionswissens (angehender) Lehrpersonen werden verschiedene Zugänge und Verfahren diskutiert, auf die im Folgenden kurz eingegangen wird.

Betont wird in Bezug auf die Erfassung professioneller Kompetenzen in der Lehrerinnen- und Lehrerbildung die Notwendigkeit, diese möglichst handlungsnah zu erfassen, weshalb der ausschließliche Einsatz von Paper-and-Pencil-Tests – insbesondere für situationsgebundenes prozedurales Wissen – kritisch gesehen wird (Aufschnaiter & Blömeke, 2010; Kunter & Pohlmann, 2015; Riese, 2009). In diesem Zusammenhang ist weiterhin umstritten, ob eher offene oder geschlossene Aufgabenformate eingesetzt werden sollten. Während bei offenen Formaten oft die geringe Reliabilität der zur Auswertung nötigen Kodierverfahren kritisiert wird, liegt deren größter Vorteil darin, „dass sie vermeiden, durch präzise Fragestellungen bereits den Erwartungshorizont vorzugeben und somit helfen einzuschätzen, welche ‚Denkansätze‘ Probanden initial bei der Bearbeitung einsetzen“ (Aufschnaiter & Blömeke, 2010, S. 364).

Aufschnaiter und Blömeke (2010) betonen für die Entwicklung von Tests außerdem die Bedeutsamkeit von Pilotierungen und Expertenvalidierungen, da generell noch wenig über das Vorwissen von Studierenden bekannt ist. Dabei erscheint es besonders wichtig, dass sich Tests explizit an evidenzbasierten Merkmalen guten Unterrichts orientieren (Aufschnaiter & Blömeke, 2010). Neben sogenannten Basisdimensionen guten Unterrichts wie einem effektiven Classroom Management oder der kognitiven Aktivierung der Lernenden ist auch das Erteilen konstruktiven Feedbacks ein Aspekt, der sich in vielen Metaanalysen als sehr lernwirksam erwiesen hat. Besondere Aufmerksamkeit für die Lehrerinnen- und Lehrerbildung hat das Merkmal Feedback vor allem durch die starke Resonanz auf die Hattie-Studie (2013) erhalten. Dieser konnte für Feedback eine sehr hohe Effektstärke von  $d = 0.73$  ermitteln. Dabei muss allerdings beachtet werden, dass Feedback nicht per se lernförderlich ist, sondern gewisse Qualitätskriterien erfüllen sollte.

## Was kennzeichnet gutes Feedback?

Nach Lipowsky (2015) wird Feedback in Anlehnung an Mory (2004) als „jede Art von Rückmeldung verstanden, die sich auf die Leistung oder das Verständnis des Lernenden bezieht, diesen über die Richtigkeit seiner Antwort bzw. seiner Aufgabenlösung informiert [...] oder ihm inhaltliche und/oder strategische Hilfen und Informationen zu seinem Bearbeitungsprozess zur Verfügung stellt“ (S. 82). Durch diese Definition wird deutlich, dass bloßes Lob ohne Bezug zur vorangegangenen Leistung der Lernenden im engeren Sinn nicht als Feedback verstanden wird (Lipowsky, 2015).

Geht es um die Qualität von Feedback, so kann natürlich „nicht generell (für alle Lehrziele, Lernereigenschaften und Unterrichtssituationen) behauptet werden, bestimmte Feedbackarten seien grundsätzlich immer anderen Formen vorzuziehen“ (Jacobs, 2002, S. 31). Geht man aber von den Funktionen aus, die Feedback erfüllen sollte, so lassen sich dennoch einige Kriterien guten Feedbacks ableiten, die sich auch in empirischen Studien als

wirksam erweisen. Auf diese Studien kann aus Platzgründen nur zusammenfassend eingegangen werden.

So hat Feedback zunächst eine informierende Funktion und sollte die Diskrepanz zwischen der momentanen Leistung und dem gewünschten Lernziel verringern (Hattie & Timperley, 2007). Hierzu führen Hattie und Timperley (2007) aus, dass Feedback die Fragen beantworten sollte, was das Lernziel ist (Feed Up), wo der Lernende aktuell steht (Feed Back) und welche konkreten Schritte als nächstes sinnvoll wären (Feed Forward). Diese Kriterien lassen sich aus verschiedenen Metaanalysen zur Effektivität von Feedback ableiten, welche starke Effekte vor allem für informatives Feedback belegen, wohingegen Lob und extrinsische Motivation geringere, teils sogar negative Effekte aufweisen (zsf. Hattie, 2013; Hattie & Timperley, 2007). Eine der umfassendsten Metaanalysen stammt von Kluger und DeNisi (1996), die u.a. zeigen, dass ein informatives Feedback insbesondere auch bei richtigen Antworten lernförderlich ist und einen Vergleich bisheriger (fehlerhafter) Lösungsversuche mit der richtigen Lösung beinhaltet. Auch hier bestätigt sich die geringe Wirksamkeit einfachen Lobs.

Rückmeldungen sollten sich zudem möglichst nicht auf die Person des Lernenden beziehen, sondern die Aufgabe, der Lösungsprozess sowie die Selbstregulation der Lernenden sollten im Vordergrund stehen. Dabei stellt ein aufgabenbezogenes, z.B. korrekatives Feedback die Grundlage dar, um den Prozess und die Selbstregulation zu fördern (Hattie & Timperley, 2007). Es weist in verschiedenen Metaanalysen sehr hohe Effektstärken von bis zu  $d = 1.13$  auf (Lysakowski & Walberg, 1982).

Obwohl elaboriertes Feedback den Lernenden hilfreiche Informationen zur Verfügung stellen sollte und wirksamer ist als einfaches Feedback (Bangert-Drowns et al., 1991; Kluger & DeNisi, 1996), sollte es im Sinne der kognitiven Aktivierung der Lernenden nicht zu viel vorwegnehmen und ihnen – insbesondere bei anspruchsvollen Aufgaben – ausreichend Bedenkzeit gewähren: „Die Kunst besteht darin, die richtige Form von Feedback auf oder knapp über dem Niveau zu geben, auf dem die Lernenden gerade arbeiten“ (Hattie, 2013, S. 210). Zusätzlich zu ihrer informierenden Funktion sollten Rückmeldungen die Lernenden auch zur Weiterarbeit ermutigen und motivieren (zsf. Richert, 2005). Dazu kann auch der Umgang mit Fehlern als Lernchance gehören (z.B. Hascher & Hagenauer, 2010).

Letztlich misst sich die Qualität von Feedback vor allem daran, ob es nützlich für die Lernenden ist. Damit Lernende von Rückmeldungen profitieren können, ist es daher auch wichtig, dass Lehrpersonen sich intensiv mit den zu bearbeitenden Aufgaben auseinandersetzen und versuchen, den Lernstand sowie die Lernfortschritte der Schülerinnen und Schüler zu analysieren, um adäquat darauf eingehen zu können (zsf. Lotz & Lipowsky, 2015).

## Selbsteinschätzungen und Testleistungen im Vergleich

Warum ist die systematische Erfassung von Wissen überhaupt nötig? Häufig wird zur Evaluation von Maßnahmen in der Lehrerinnen- und Lehrerbildung schließlich auf Selbsteinschätzungen zurückgegriffen (zsf. Cramer, 2010; König, Kaiser & Felbrich, 2012), die in der Erhebung und Auswertung deutlich weniger aufwändig sind. Dabei werden u.a. Selbstwirksamkeitserwartungen von (angehenden) Lehrpersonen erfragt, also das Ausmaß ihrer „subjektive(n) Gewissheit, neue oder schwierige Anforderungssituationen auf Grund

eigener Kompetenz bewältigen zu können“ (Schwarzer & Jerusalem, 2002, S. 35). Aufgrund der Subjektivität und der damit vermuteten geringen Verlässlichkeit von Einschätzungen eigener Kompetenzen werden aber im Rahmen einer evidenzbasierten Lehrerinnen- und Lehrerbildung Evaluationsmaßnahmen gefordert, die den Lernzuwachs mit Testverfahren erfassen (König et al., 2012). Bisher haben sich nur wenige Studien mit der Frage beschäftigt, inwieweit Selbsteinschätzungen und mit Testverfahren erhobene Leistungen von (angehenden) Lehrpersonen zusammenhängen (vgl. zsf. König et al., 2012).

Im Rahmen der LEK-Studie wurde der Zusammenhang zwischen der selbsteingeschätzten Handlungskompetenz und dem pädagogischen Professionswissen von angehenden Lehrpersonen untersucht. Die resultierenden Korrelationen sind zum Teil zwar positiv und bedeutsam, aber sehr gering. In einigen Teilfacetten zeigen sich gar keine Zusammenhänge (König & Tachtsoglou, 2012). Selbsteinschätzung und Test erfassen hier allerdings nicht die exakt gleichen Aspekte, da es um die Einschätzung der eigenen *Handlungskompetenz* ging, während der Test das *pädagogische Wissen* erfassen sollte, welches zwar als bedeutsam für die Handlungskompetenz angesehen wird, aber nicht damit gleichgesetzt werden kann.

Auch anhand von Daten aus der TEDS-M-Studie zeigen König und Kollegen (2012) nur geringe Zusammenhänge zwischen Wissen und Selbsteinschätzung angehender Lehrpersonen verschiedener Schularten in Deutschland und den USA. Auch hier sind die Korrelationen lediglich für gewisse Teilstichproben statistisch signifikant. Während das pädagogische Wissen anhand eines Tests gemessen wurde, wurde zur Erfassung der Selbsteinschätzung die „Überzeugung, von der Lehramtsausbildung auf das Unterrichten vorbereitet zu sein“ (König et al., 2012, S. 478) mit einem Fragebogen erfasst.

Schulte, Bögeholz und Watermann (2008) untersuchten den Zusammenhang zwischen Testwerten zum pädagogischen Professionswissen und Selbstwirksamkeitserwartungen bei angehenden Lehrkräften in unterschiedlichen Stadien der Ausbildung. Auch hier resultieren signifikante Korrelationen verschiedener Facetten, die aber – wie auch in den anderen Studien – nur gering ausfallen.

Für die geringer als erwartet ausfallenden Zusammenhänge könnten folgende Erklärungen in Frage kommen: Zum einen ist denkbar, dass Studierende ihre Kompetenzen tatsächlich nicht adäquat einschätzen können, da sie – beispielsweise aufgrund noch geringer (Praxis-)Erfahrungen – eventuell einer Kompetenzillusion (Koriat & Bjork, 2005) unterliegen. Während Tests meist sehr konkrete Fragen zu bestimmten Inhalten beinhalten, stellen Selbsteinschätzungen außerdem eher „holistische Abbildungen“ (König & Tachtsoglou, 2012, S. 296) dar und werden auch von allgemeinen Persönlichkeitsmerkmalen beeinflusst (Cramer, 2010). Des Weiteren könnten die geringen Zusammenhänge auch durch „(m)otivationale Voreingenommenheit, soziale Erwünschtheit, aber auch das Fehlen eines gemeinsamen Vergleichsmaßstabes der Zielgruppe“ erklärt werden (König et al., 2012, S. 488). Zusammenfassend lässt sich aus den bisherigen Studien ableiten, dass „Wissens-tests nicht durch subjektive Kompetenzeinschätzungen ersetzt werden können und vice versa“ (König & Tachtsoglou, 2012, S. 296) und „dass wir bislang nur wenig darüber wissen, was wir mit Selbsteinschätzungen tatsächlich messen“ (König et al., 2008, S. 488).

## Hypothesen

Im Vordergrund dieses Beitrags stehen die Vorstellung und Überprüfung der Eignung eines Instruments zur Erfassung konzeptuellen Wissens von Lehramtsstudierenden zum Thema Feedback. Gefragt wird in diesem Zuge auch nach Zusammenhängen des durch den Test erfassten Wissens mit den Selbsteinschätzungen des Wissens(zuwachses) durch die Studierenden.

Zunächst wird überprüft, ob das konzeptuelle Wissen von Lehramtsstudierenden zum Thema Feedback mit einem selbst entwickelten Instrument anhand von offenen Fragen objektiv und reliabel erfasst werden kann:

*Frage 1: Lässt sich das konzeptuelle Wissen über Feedback mit dem entwickelten Instrument objektiv und reliabel erfassen?*

Darauf aufbauend wird zunächst beschrieben, wie sich mittels des entwickelten Instruments das Wissen und der Wissenszuwachs von Studierenden beschreiben lassen. Dabei wird auch überprüft, ob eine kriteriale Auswertung und eine Globaleinschätzung der Qualität der Studierendenantworten zusammenhängen und sich damit wechselseitig validieren:

*Frage 2: Wie lassen sich das Wissen und der Wissenszuwachs der Studierenden mit den Auswertungen durch das Instrument beschreiben?*

Zuletzt wird analysiert, ob sich bedeutsame Zusammenhänge zwischen den mit dem Test erfassten Wissen und der Selbsteinschätzung des Wissens(zuwachses) durch die Studierenden zeigen:

*Frage 3: Wie hoch sind die Zusammenhänge zwischen dem mit dem Fragebogen erfassten Wissen und den Selbsteinschätzungen der Studierenden zu ihrem Wissen zu beiden Messzeitpunkten sowie zu ihrem Wissenszuwachs durch das Seminar?*

Diese Frage ist vor allem vor dem Hintergrund relevant, dass zur Evaluation universitärer Seminare oftmals Selbsteinschätzungen des Wissenszuwachses von Studierenden genutzt werden. Sollten sich geringe Zusammenhänge zwischen den Selbsteinschätzungen und dem erfassten Wissen zeigen, würde dies die Bedeutsamkeit der systematischen Überprüfung des Wissenszuwachses unterstreichen. Zeigen sich hingegen hohe Zusammenhänge, würde dies auf die Verlässlichkeit von Selbsteinschätzungen hinweisen, deren Erhebung deutlich weniger aufwändig ist als die Testung.

## Datengrundlage und methodisches Vorgehen

Das konzeptuelle Wissen der Lehramtsstudierenden wurde im Rahmen von drei Seminaren erhoben, die an der Universität Kassel im erziehungswissenschaftlichen Kernstudium angeboten wurden. Die Seminare fanden im Wintersemester 2015/16 und im Sommersemester 2016 in Form von jeweils dreitägigen Blockseminaren statt und trugen den Titel „Gutes Feedback ist mehr als ‚toll gemacht!‘ – Wie gibt man Schülerinnen und Schülern hilfreiche Rückmeldungen?“.

In den Seminaren wurden Qualitätsmerkmale guten Feedbacks – insbesondere anhand des Feedback-Modells von Hattie und Timperley (2007) – erarbeitet und von den Studierenden bei der Analyse von Videos angewendet. Da die genauere Gestaltung der Semina-

re für den vorliegenden Beitrag nicht relevant ist, wird sie hier nicht weiter erläutert (vgl. hierzu Hess, 2017).

Der Lernzuwachs der Studierenden wurde in einem Prä-Post-Design auf verschiedenen Ebenen überprüft. So wird neben dem in diesem Beitrag im Fokus stehenden konzeptuellen Wissen auch die professionelle Wahrnehmung der Studierenden von Feedback-Situationen sowie ihre Handlungskompetenz im Bereich Feedback untersucht. Auf diese beiden Aspekte wird im vorliegenden Beitrag nicht näher eingegangen.

### *Stichprobe*

An den drei Seminaren nahmen insgesamt 54 Studierende teil, wovon 87 % weiblich waren. Die Studierenden waren zum Zeitpunkt des Seminars im Schnitt 24 Jahre alt und im sechsten Semester. Etwas mehr als die Hälfte der Studierenden hatte bereits eine Veranstaltung besucht, in der das Thema Feedback in einer oder mehreren Sitzungen behandelt wurde. Eine gesamte Veranstaltung zum Thema hatte aber noch keiner/keine der Studierenden besucht. Die meisten der teilnehmenden Studierenden waren für das Lehramt an Gymnasien sowie Haupt- und Realschulen eingeschrieben, gefolgt vom Lehramt an Grundschulen und vom Lehramt für berufliche Schulen (vgl. Tabelle 1).

Anzahl der Studierenden		54
Anteil weibliche Studierende		87.04 %
Lehramt	L1 = Lehramt an Grundschulen	20.37 %
	L2 = Lehramt an Haupt- und Realschulen	35.19 %
	L3 = Lehramt an Gymnasien	37.04 %
	L4 = Berufs- und Wirtschaftspädagogik	7.41 %
Fachsemester (Mittelwert)		5.96
Alter (Mittelwert)		24.22
Anteil der Studierenden, die das Thema Feedback bereits in anderen Veranstaltungen behandelt haben		53.70 %

Tab. 1: Stichprobe.

Für alle in Tabelle 1 aufgeführten Variablen wurde überprüft, ob sie systematisch mit der Höhe des konzeptuellen Wissens im Vortest zusammenhängen. Hierbei zeigten sich keinerlei signifikante Zusammenhänge.

### *Instrument: Wissenstest*

Der Test zum konzeptuellen Wissen wurde sowohl vor als auch unmittelbar nach dem Seminar eingesetzt. Die Studierenden bekamen hierzu zwei Aufgaben gestellt, die offen beantwortet werden sollten: „(1) Definieren Sie kurz, was Sie unter ‚Feedback‘ von der Lehrperson an Schülerinnen und Schüler verstehen. (2) Nennen Sie Kriterien, die für Sie ‚gutes Feedback‘ von der Lehrperson an Schülerinnen und Schüler beschreiben.“ Die Studierenden hatten bis zu 15 Minuten Zeit, die Fragen zu beantworten, wobei die durchschnittliche Bearbeitungszeit bei etwa sechs Minuten lag.

Für die Auswertung wurde einerseits ein kriteriales, mittel inferentes Verfahren entwickelt und zum anderen ein globales, hoch inferentes Schätzverfahren. Der Grad an Inferenz beschreibt das Ausmaß an Schlussfolgerungen, das zur Beurteilung nötig ist und kann als Kontinuum verstanden werden (vgl. Lotz, Berner & Gabriel, 2013). Beide Verfahren wurden auf Basis des Forschungsstands zu effektivem Feedback deduktiv entwickelt und in-

duktiv durch die Analyse von Studierendenantworten aus einer Pilotierungsstichprobe ergänzt.

Für die mittel inferente Auswertung wurden insgesamt 11 Kriterien guten Feedbacks festgelegt, für die jeweils analysiert wurde, inwiefern sie von den Studierenden nicht (0 Punkte), im Ansatz (1 Punkt) oder vollständig richtig (2 Punkte) genannt wurden. Für jedes Merkmal wurde ausführlich anhand von Kodierregeln und Beispielen beschrieben, welche Formulierungen der Studierenden welchem Wert entsprechen. Dabei wurde nicht danach unterschieden, ob sich die Formulierungen auf die Aufgabe 1 oder 2 bezogen, d.h. die Antworten der Studierenden wurden für die Auswertung zusammen betrachtet. Tabelle 2 stellt die 11 Items im Überblick dar.

Item	Diese Kategorie ist enthalten, wenn beschrieben wird, dass ...
Feed-Up	... Feedback eine Diskrepanz zwischen der aktuellen Leistung und dem Lernziel aufzeigen soll (z.B. „gutes Feedback sollte dem Schüler das Ziel verdeutlichen“).
Feed-Back	... Feedback eine Rückmeldung zu bereits Erreichtem ist und dass es möglichst informativ/begründet/detailliert o.ä. sein sollte (z.B. „ausführliche Rückmeldung bezüglich des Arbeits- und Lernverhaltens sowie des Leistungsstands“).
Feed-Forward	... gutes Feedback den Lernenden Tipps, Hilfen, Impulse, Verbesserungsvorschläge o.ä. gibt (z.B. „sollte Anregungen und Fragen enthalten, die das Kind zu den nächsten Schritten bringt, damit es das Ziel erlangt“).
Selbstständigkeit	... die Lehrperson den Lernenden nicht zu viel vorgeben sollte (z.B. „der Lehrer sollte dem Schüler die Lösung nicht vorsagen und ihm eigene Denkmöglichkeiten lassen“).
Wartezeit	... beim Erteilen von Feedback das Timing eine Rolle spielt und die Lehrperson den Lernenden ausreichend Zeit zum Nachdenken lassen sollte (z.B. „genug Bedenkzeit geben“).
Motivation/ Ermutigung	... Feedback für die Lernenden motivierend sein und sie in ihrem Lernprozess ermutigen sollte (z.B. „motivierend, z.B. durch individuelle Bezugsnorm“).
Diagnostik/ Analyse der Aufgabe	... die Lehrperson sich genau mit der Aufgabe und/oder den Lösungsversuchen der Lernenden beschäftigt, indem beispielsweise beobachtet oder Leistung dokumentiert wird, die Aufgabe gut vorbereitet oder mögliche Fehler antizipiert werden (z.B. „intensiv mit der Arbeit der Schüler auseinandersetzen und beobachten“).
Nutzen für den Schüler	... gutes Feedback sich dadurch auszeichnet, dass es den Lernenden hilft und diese danach besser vorankommen als vorher (z.B. „das Feedback muss dem Schüler auch wirklich weiterhelfen“).
Vermeidung des Personenbezugs	... gutes Feedback sich nicht auf persönliche Merkmale – wie die Intelligenz der Lernenden – bezieht (z.B. „es sollte nicht zur Kritik an der Person, sondern als Kritik an erbrachten Leistungen genutzt werden“).
Förderung der Selbstregulation	... gutes Feedback die Selbstregulation der Lernenden fördert (z.B. „etwas wodurch Schüler die Chance bekommen, ihr eigenes Lernverhalten zu beobachten, zu verändern, zu hinterfragen“).
Fehler als Lernchance	... Fehler als Lernchance verstanden werden sollten und mit ihnen produktiv umgegangen werden sollte (z.B. „Feedback zeigt dem Schüler, dass Fehler nicht schlimm, sondern sogar gut sind und gemacht werden dürfen“).

Tab. 2: Überblick über die 11 mittel inferent erfassten Merkmale guten Feedbacks und Ankerbeispiele.

Für das globale hoch inferente Rating wurde die gesamte Antwort der Studierenden herangezogen und auf einer sechsstufigen Skala in ihrer Qualität eingeschätzt („1 = sehr schlecht“; „2 = eher schlecht“; „3 = befriedigend“; „4 = gut“; „5 = sehr gut“; „6 = ausgezeichnet“). Dabei sollte beurteilt werden, ob die Studierenden ein umfassendes Bild der relevanten Merkmale guten Feedbacks haben, ob die Aussagen inhaltlich korrekt und

ausführlich sind und ob die genannten Merkmale tatsächlich relevant für gutes Feedback sind. Jede der sechs Ratingstufen wurde dabei kurz verbal beschrieben und mit Ankerbeispielen veranschaulicht. Die „6 = ausgezeichnet“ wurde beispielsweise wie folgt definiert: „Eine ‚6‘ wird vergeben, wenn nahezu alle Kategorien in der Definition enthalten sind. Die Definition ist sehr ausführlich und weist keine inhaltlichen Fehler auf. Es werden ausschließlich relevante Merkmale beschrieben.“ Abbildung 1 verdeutlicht jeweils anhand eines Beispiels aus dem Vor- und Nachtest das Vorgehen bei der Auswertung.

<p><b>Beispiel aus dem Vortest: Definition von Feedback und Kriterien guten Feedbacks eines Studierenden mit Bepunktung</b></p> <p>Stärken und Schwächen der SuS zeigen.</p> <ul style="list-style-type: none"> <li>• Empathie, individuell auf Schüler eingehen → DIAGNOSTIK/ANALYSE DER AUFGABE (1)</li> <li>• Verbesserungsmöglichkeiten geben → FEED-FORWARD (2)</li> </ul> <p>MITTEL INFERENTE KODIERUNG: 2 von 11 möglichen Merkmalen genannt; 3 von 22 möglichen Punkten erreicht</p> <p>HOCH INFERENTES RATING: Ratingstufe 1 („sehr schlecht“)</p>
<p><b>Beispiel aus dem Nachtest: Definition von Feedback und Kriterien guten Feedbacks eines Studierenden mit Bepunktung</b></p> <p>Bei Feedback erhält der Lernende Informationen über die Bearbeitung einer Aufgabe. Es wird also genau darauf eingegangen, wie richtig die Lösung des Schülers ist → FEED-BACK (2). Außerdem sollten Tipps, Impulse und Hilfestellungen an den Schüler herangetragen werden, um dessen Lernen im Allgemeinen zu verbessern und zu konkretisieren. → FEED-FORWARD (2)</p> <ul style="list-style-type: none"> <li>• Motivation des Schülers; Individuelle Vergleiche („Das klappt jetzt schon viel besser“); Keine sozialen Vergleiche (mit anderen Kindern/ Leistungen vergleichen) → MOTIVATION/ERMUTIGUNG (2)</li> <li>• Prozess- und produktbezogen; Sachlich; Kein personenbezogenes Feedback → VERMEIDUNG DES PERSONENBEZUGS (2)</li> <li>• Schülerbeobachtung vorher → DIAGNOSTIK/ANALYSE DER AUFGABE (2)</li> <li>• Für das Kind verständlich → NUTZEN (1)</li> <li>• genug Zeit zum Denken lassen → WARTEZEIT (2)</li> </ul> <p>MITTEL INFERENTE KODIERUNG: 7 von 11 möglichen Merkmalen genannt; 13 von 22 möglichen Punkten erreicht</p> <p>HOCH INFERENTES RATING: Ratingstufe 4 („gut“)</p> <p>(1) = Merkmal im Ansatz genannt; (2) = Merkmal vollständig richtig genannt</p>

Abb. 1: Auswertungsbeispiel.

### *Vorgehen bei der Auswertung des Wissenstests*

Die Auswertungen wurden von der Entwicklerin des Beurteilungssystems und einer geschulten Kodiererin (Lehramtsstudentin) durchgeführt. Die Schulung umfasste zwei Tage und enthielt neben der Erläuterung der Kriterien auch Übungskodierungen an der Pilotierungsstichprobe. Im Anschluss an die Schulung wurde das gesamte Material von beiden Kodiererinnen unabhängig voneinander ausgewertet, wobei jeweils nach etwa fünfzehn Fragebögen eine Prüfung der Übereinstimmungen und Reliabilitäten stattfand. Für weitere Analysen mit den Daten wurde bei Nicht-Übereinstimmungen in den mittel inferenten Kodierungen ein Konsensurteil gebildet, für das hoch inferente Rating der Mittelwert beider Kodiererinnen.

### *Fragen zur Selbsteinschätzung des Wissens(zuwachses) durch die Studierenden*

Zur Erfassung des selbsteingeschätzten Wissens wurde im Vortest sowie im Nachtest jeweils eine Einschätzung zur gleichen Aussage verlangt: „Ich weiß, welche Kriterien ein gutes Feedback ausmachen.“ Im Nachtest wurden die Studierenden zusätzlich um eine Einschätzung zu folgendem Item gebeten: „Im Bereich des theoretischen Wissens zum Thema Feedback habe ich durch das Seminar viel dazugelernt“. Alle Items wurden auf

einer vierstufigen Skala von „1 = trifft überhaupt nicht zu“ bis „4 = trifft voll zu“ von allen 54 Studierenden jeweils nach der Bearbeitung des Wissenstests ausgefüllt.

### *Statistische Analysen*

Um die Objektivität und Reliabilität der Auswertung zu prüfen, wurden für die mittel inferenten Kodierungen die prozentualen Übereinstimmungen zwischen den beiden Kodierenden berechnet. Zusätzlich wird Cohens Kappa berichtet, ein um die zufällige Übereinstimmung bereinigter Wert. Bei der prozentualen Übereinstimmung wurden Werte von  $PÜ \geq 85\%$  als Mindestwert festgelegt, Cohens Kappa sollte  $\kappa \geq .70$  betragen. Für die hoch inferenten Ratings wurde mit dem GT-Programm für Generalisierbarkeitsstudien (Ysewijn, 1996) der Generalisierbarkeitskoeffizient berechnet. Dieser entspricht dem Reliabilitätskoeffizienten aus der klassischen Testtheorie. Der relative G-Koeffizient vergleicht die Rangfolgen in der Beurteilung der beiden Kodierenden, während der absolute G-Koeffizient auch die Höhe der Beurteilung mitberücksichtigt. Außerdem berechnet das GT-Programm die Anteile der Varianz, die auf Unterschiede zwischen den Ratern, tatsächliche Unterschiede in den Tests oder die Interaktion zwischen Test und Rater sowie unsystematische Fehler zurückgehen. Für eine reliable Auswertung sollte der G-Koeffizient bei  $g \geq .70$  liegen und der größte Varianzanteil sollte auf Unterschiede zwischen den Tests zurückzuführen sein (für eine ausführliche Beschreibung der Kennwerte zur Prüfung der Reliabilität vgl. Lotz et al., 2013).

Um das Wissen der Studierenden mit Hilfe des Instruments zu beschreiben (*Fragestellung 2*) und die Zusammenhänge zur Selbsteinschätzung zu untersuchen (*Fragestellung 3*), werden Häufigkeiten, deskriptive Statistiken und – aufgrund der Nicht-Normalverteilung der Daten – Korrelationen nach Spearman-Rho berichtet. Zudem wurden einfaktorische Varianzanalysen mit Messwiederholung berechnet. Da die Stichprobe mehr als 30 Fälle pro Messzeitpunkt umfasst, ist diese Analyse trotz nicht normalverteilter Daten als robust einzuschätzen (Bortz & Schuster, 2010).

## **Ergebnisse**

### *Ergebnisse zur Objektivität und Reliabilität des Instruments (Fragestellung 1)*

Für jedes der 11 mittel inferent erfassten Items wurde überprüft, wie häufig beide Kodierenden den exakt gleichen Wert erteilt haben (0 = Merkmal nicht genannt; 1 = Merkmal ansatzweise genannt; 2 = Merkmal vollständig genannt). Die prozentualen Übereinstimmungen (PÜ) liegen zwischen  $PÜ = 90.74\%$  (Feed-Back) und  $PÜ = 99.07\%$  (Fehler als Lernchance). Die Werte für Cohens Kappa bewegen sich zwischen  $\kappa = .86$  und  $\kappa = .97$ . Damit werden die geforderten Mindestwerte von  $PÜ \geq 85\%$  und  $\kappa \geq .70$  übertroffen, sodass von einer ausreichend hohen Objektivität ausgegangen werden kann.

Für das hoch inferente Rating wurden der relative ( $g_{\text{relativ}}$ ) und der absolute G-Koeffizient ( $g_{\text{absolut}}$ ) sowie die Varianzkomponenten berechnet. Die Werte liegen mit  $g_{\text{relativ}} = .94$  sowie  $g_{\text{absolut}} = .94$  deutlich über dem Mindestwert von  $g \geq .70$ , sodass von reliablen Einschätzungen ausgegangen werden kann. Mit 89 % geht der größte Anteil der Varianz auf tatsächliche Unterschiede zwischen den Tests zurück, wohingegen sich 0% der Varianz durch die Personen der Raterinnen erklären lassen. 11% der Unterschiede in den Auswertungen lassen sich auf die Interaktion zwischen Test und Raterin sowie auf unsystematische Fehler zurückführen.

Damit kann sowohl für die kriteriale, mittel inferente Auswertung als auch für das globale, hoch inferente Rating eine zufriedenstellende Objektivität und Reliabilität der Datenauswertung bestätigt werden.

### *Ergebnisse zur Beschreibung des Wissens und des Wissenszuwachses (Fragestellung 2)*

Aus den mittel inferent erfassten Ergebnissen können zwei Kennwerte gebildet werden. Für den ersten Kennwert („Breite und Qualität des Wissens“) wurden die Punkte für die 11 Kriterien zu einem Gesamtscore aufsummiert. Da für jedes Merkmal 0, 1 oder 2 Punkte erzielt werden konnten, können hier maximal 22 Punkte erreicht werden, wenn alle 11 Kriterien vollständig genannt wurden. Ein Wert von 4 Punkten könnte hier entweder zustande kommen, wenn ein Studierender beispielsweise vier Merkmale ansatzweise nennt (4x1 Punkt), aber auch wenn er jeweils zwei Merkmale vollständig (2x2 Punkte) beschrieben hat. Für den zweiten, etwas weniger differenzierten Kennwert wurde nicht zwischen ansatzweiser und vollständiger Nennung unterschieden, sodass das theoretische Maximum bei 11 liegt. Der zweite Wert gibt demnach also an, wie viele der 11 Kriterien von den Studierenden genannt wurden („Breite des Wissens“). Beim hoch inferenten Rating bedeutet die 1 eine sehr geringe Qualität und der Wert 6 eine ausgezeichnete Qualität des durch den Test erfassten konzeptuellen Wissens. Bei allen drei Kennwerten zeigen hohe Werte also ein hohes Wissen über Feedback an.

	Breite und Qualität des Wissens (mittel inferent kodiert, 0-22 Punkte)	Breite des Wissens (mittel inferent kodiert, 0-11 Punkte)
	gesamt/Prätest/Posttest	gesamt/Prätest/Posttest
<b>Qualität des Wissens (hoch inferent eingeschätzt, 1-6)</b>	.91**/.84**/.89**	.86**/.73**/.85**
<b>Breite und Qualität des Wissens</b>		.95**/.88**/.95**

\*\* Die Korrelation ist auf dem 0.01 Niveau signifikant (zweiseitig).

Tab. 3: Korrelationen (Spearman-Rho) der unterschiedlichen Kennwerte.

Tabelle 3 enthält die Korrelationen dieser drei Kennwerte untereinander, wobei der erste Korrelationskoeffizient den Wert für die gesamte Stichprobe von  $N = 108$  Tests über beide Messzeitpunkte hinweg abbildet, der zweite Koeffizient die Zusammenhänge der jeweiligen Daten aus dem Prätest repräsentiert ( $n = 54$ ) und der dritte Koeffizient die entsprechenden Zusammenhänge innerhalb des Posttests ( $n = 54$ ) angibt.

Über beide Messzeitpunkte hinweg liegen die Korrelationen der Kennwerte untereinander zwischen  $r = .86$  und  $r = .95$ . Ähnlich hohe Werte ergeben sich auch für die Korrelationen innerhalb des Posttests. Für den Prätest liegen die Werte etwas niedriger und bewegen sich zwischen  $r = .73$  und  $r = .88$ . Insgesamt sind die Zusammenhänge zwischen den beiden Kennwerten, die aus der mittel inferenten Kodierung generiert wurden, etwas höher als zwischen mittel und hoch inferenter Auswertung. Hier korreliert die „Breite und Qualität des Wissens“ etwas enger mit dem hoch inferenten Rating. Grundsätzlich zeigen die hohen, signifikanten Korrelationen aber für alle drei Kennwerte, dass sie grundsätzlich sehr Ähnliches erfassen.

Tabelle 4 stellt die deskriptiven Statistiken für alle drei Kennwerte im Prä- und Posttest dar. Sowohl die Werte für Minimum und Maximum als auch für den Mittelwert verdeutlichen für alle Kennwerte eine deutliche Verbesserung vom Vor- zum Nachtest. Während

im Prätest durchschnittlich drei Kriterien guten Feedbacks von den Studierenden genannt wurden, waren es im Posttest sechs Merkmale. Dass die Standardabweichungen von Prä- zu Posttest größer werden, zeigt an, dass die Leistungen der Studierenden nach den Seminaren stärker streuen als vorher. Trotz dieser Verbesserungen sind die Mittelwerte auch im Nachtest aber noch deutlich vom theoretischen Maximum entfernt. Beim hoch inferenten Rating werden von den sechs möglichen Punkten im Schnitt 3.44 erreicht, was einer guten bis befriedigenden Qualität der Studierendenantworten entspricht.

	MZP	Min	Max	M	SD	df	F	p	$\eta^2$
Mittel inferente Kodierung (Breite und Qualität des Wissens)	Prätest	0	10	4.61	2.37	53	56.88	.000	0.52
	Posttest	2	18	9.31	4.26				
Mittel inferente Kodierung (Breite des Wissens)	Prätest	0	7	3.30	1.49	53	59.04	.000	0.53
	Posttest	2	11	5.87	2.27				
Hoch inferentes Rating	Prätest	1.0	4.5	2.31	0.76	53	62.87	.000	0.54
	Posttest	1.0	6.0	3.44	1.06				

Tab. 4: Wissenszuwachs von Prä- und Posttest ( $N = 54$ ).

Varianzanalysen mit Messwiederholung ergeben für alle Kennwerte einen signifikanten Effekt des Messzeitpunkts, also einen bedeutsamen Wissenszuwachs der Studierenden von Prä- zu Posttest. Auch die Effektstärken von  $\eta^2 \geq 0.52$ , die sich für die drei berechneten Kennwerte kaum unterscheiden, weisen darauf hin, dass das Instrument in der Lage ist, einen Wissenszuwachs abzubilden.

Zusätzlich zu den in Tabelle 4 dargestellten deskriptiven Ergebnissen veranschaulicht Abbildung 2, welche Kriterien guten Feedbacks von den Studierenden im Vor- und Nachtest genannt wurden. Dabei wird deutlich, dass für jedes einzelne Item die Häufigkeit der Nennung vom Vor- zum Nachtest zunimmt. Allerdings ergeben sich deutliche Unterschiede in der Häufigkeit der Nennungen zwischen den Merkmalen. Während beispielsweise fast alle Studierenden im Nachtest Feed-Back und Feed-Forward als bedeutsame Merkmale guten Feedbacks benennen, gehen nur sehr wenige darauf ein, dass Fehler von der Lehrperson als Lernchance betrachtet werden sollten.

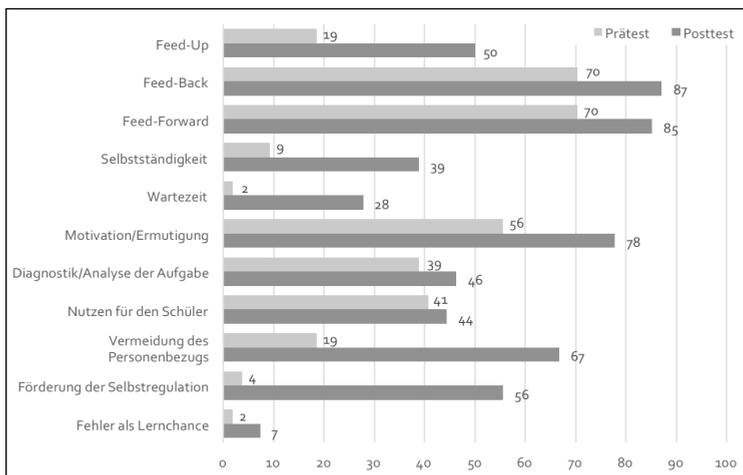


Abb. 2: Prozentualer Anteil der Studierenden, die die einzelnen Kriterien im Test ansatzweise oder vollständig nennen.