

Puffer, Gabriele

Messung professioneller Kompetenzen von Studierenden zur Evaluation einer fachdidaktischen Lehrveranstaltung

Musikpädagogik im Spannungsfeld von Reflexion und Intervention. Münster ; New York : Waxmann 2021, S. 91-112. - (Musikpädagogische Forschung; 41)



Quellenangabe/ Reference:

Puffer, Gabriele: Messung professioneller Kompetenzen von Studierenden zur Evaluation einer fachdidaktischen Lehrveranstaltung - In: Musikpädagogik im Spannungsfeld von Reflexion und Intervention. Münster ; New York : Waxmann 2021, S. 91-112 - URN: urn:nbn:de:0111-pedocs-243348 - DOI: 10.25656/01:24334

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-243348>

<https://doi.org/10.25656/01:24334>

in Kooperation mit / in cooperation with:



WAXMANN
www.waxmann.com

<http://www.waxmann.com>

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

Johannes Hasselhorn, Oliver Kautny, Friedrich Platz (Hrsg.)

MUSIKPÄDAGOGIK IM SPANNUNGSFELD VON REFLEXION UND INTERVENTION

MUSIC EDUCATION BETWEEN
(SELF-)REFLECTIONS
AND INTERVENTIONS

Musikpädagogische Forschung

Research in Music Education

Herausgegeben vom Arbeitskreis
Musikpädagogische Forschung e. V. (AMPF)

Band 41

Proceedings of the 41st Annual Conference of the
German Association for Research in Music Education

Johannes Hasselhorn, Oliver Kautny,
Friedrich Platz (Hrsg.)

Musikpädagogik im Spannungsfeld von Reflexion und Intervention

Music Education between
(Self-)Reflections and Interventions



Waxmann 2021
Münster • New York

Bibliografische Informationen der Deutschen Nationalbibliothek

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

**Musikpädagogische Forschung, Band 41
Research in Music Education, vol. 41**

ISSN 0937-3993

Print-ISBN 978-3-8309-4272-6

E-Book-ISBN 978-3-8309-9272-1

© Waxmann Verlag GmbH, Münster 2021
Steinfurter Straße 555, 48159 Münster

www.waxmann.com
info@waxmann.com

Umschlaggestaltung: Anne Breitenbach, Münster
Satz: Roger Stoddart, Münster

Inhalt

Vorwort.....	9
--------------	---

Liste der Reviewerinnen und Reviewer	13
--	----

Andreas Lehmann-Wermser

„... es kömmt drauf an ...“	15
-----------------------------------	----

Marlon G. Schneider

Concept Maps

Erprobung und Reflexion eines Diagnoseinstruments zur Erfassung von Lernvoraussetzungen und -prozessen im Musikunterricht	33
--	----

Concept Maps

*The Testing of and Reflection upon a Diagnostic Tool for Recording Learning
Prerequisites and Processes in Music Lessons*

Georg Brunner, Thade Buchborn, Bernd Clausen,

Werner Jank & Silke Schmid

Change Management im Lehramtsstudium –

Kohärenz und Professionsorientierung	53
--	----

*Change Management Processes in Music Teacher Training:
Coherency and Professional Orientation*

Georg Brunner

Inverted Classroom Model (ICM) und E-Lecture als Konzept für eine musikbezogene Hochschullehre: eine explorative Vergleichsstudie	73
--	----

*The Inverted Classroom Model (ICM) and e-Lectures in Music-related
University Teaching: An Exploratory Comparative Study*

Gabriele Puffer

Messung professioneller Kompetenzen von Studierenden zur Evaluation einer fachdidaktischen Lehrveranstaltung	91
---	----

*Assessing Future Music Teachers' Professional Competencies for the
Evaluation of a Teacher Training Program*

Isolde Malmberg

Musizieren als Entwicklungsaufgabe im Schulpraktikum –
erste Ergebnisse der qualitativen Erhebung TRANSFER zu Mentor*in-
Mentee-Dynamiken beim Klassenmusizieren in schulischen Praxisphasen. 113

*Music Making as a Developmental Task in School Internships –
First Results of the Qualitative Survey TRANSFER on Mentor-
Mentee Dynamics in Practical School Phases*

Anna-Lisa Jeismann & Ulrike Kranefeld

(Un-)Eindeutige Anregungen
Zur Rekonstruktion von Handlungsmustern bei der Begleitung
von Prozessen des Musik-Erfindens 135
*(Un)Specific Advices: Reconstructing Interactional Patterns
of Instructors in School Composition Settings*

*Linus Eusterbrock, Marc Godau, Matthias Haenisch,
Matthias Krebs & Christian Rolle*

Von ‚inspirierenden Orten‘ und ‚Safe Places‘: die ästhetische
Nutzung von Orten in der Appmusikpraxis 155
*On ‘Inspiring’ and ‘Safe Places’: The Aesthetic Use of Place in
Music-App Practices*

Peter W. Schatt

„geschichtlich durch und durch“?
Überlegungen zur musikpädagogischen Relevanz einer
kulturwissenschaftlichen Perspektive auf das ‚musikalische Material‘ 173
*“geschichtlich durch und durch”? Considerations on Music Pedagogical
Relevance of a Cultural Studies Perspective on the ‘Musical Material’*

Hans-Ulrich Schäfer-Lembeck

Lehre des Musikalisch-Künstlerischen –
professionell, als Profession? 191
Teaching the ‘Musical Artist’ – Professionally, or as a profession?

<i>Lina Oravec & Julia Steffens unter Mitarbeit von Susanne Becker</i>	
<i>„Solmisieren?! Manchmal hilft's, manchmal ist es ein Zeitverschwender.“</i>	
Erfolge und Grenzen relativer Solmisation in der Grundschule – eine Interviewstudie nach der Grounded Theory	213
<i>„Solfège?! Sometimes helpful, sometimes a waste of time.“ Outcomes and Limitations of the Movable-Do Solfège Approach in German Elementary Schools – a Grounded Theory Study</i>	
 <i>Olivier Blanchard</i>	
<i>„Die Songs der Schüler*innen“ und „die Songs der Lehrer*innen“</i>	
Die Herstellung kultureller Differenzen im Musikunterricht	233
<i>“Students’ Songs” and “Teachers’ Songs”: Cultural Differences in Music Lessons</i>	
 <i>Charlotte Lietzmann</i>	
<i>Die Kunst der Distinktion – Prozesse von Ein- und Ausschluss</i>	
im Feld musikpädagogischer Studiengänge	251
<i>The Art of Distinction – Processes of In- and Exclusion in the Field of Higher Music Education</i>	
 <i>Jürg Huber und Christoph Marty</i>	
<i>Die diskursive Behauptung einer eigenen Musikpädagogik in der Deutschschweiz im Spiegel von Rezensionen.</i>	
	277
<i>The Discursive Assertion of an Individual Music Education in German-Speaking Switzerland as Reflected in Reviews</i>	
 <i>Stefan Orgass</i>	
<i>Gegenstandsbereiche musikpädagogischer Forschung in reflexionslogischer Rekonstruktion.</i>	
	297
<i>Topics of Research in Music Pedagogy According to the Logics of Reflection</i>	

Gabriele Puffer

Messung professioneller Kompetenzen von Studierenden zur Evaluation einer fachdidaktischen Lehrveranstaltung

Assessing Future Music Teachers' Professional Competencies for the Evaluation of a Teacher Training Program

Wie lassen sich künftige Musiklehrkräfte angemessen auf die Herausforderungen vorbereiten, die beim Singen mit heterogenen Gruppen im Musikunterricht auf sie zukommen? Wie können entsprechende Lernergebnisse dokumentiert werden?

Im Rahmen der Qualitätsoffensive Lehrerbildung des BMBF entwickelte der Lehrstuhl für Musikpädagogik der Universität Augsburg ein Lehrformat, mit dessen Hilfe domänenspezifisches Professionswissen und situationsspezifische Fähigkeiten von Lehramtsstudierenden gefördert werden sollen. Ein wesentliches methodisches Element des Seminars ist fallbasiertes Arbeiten mit Video- und Audiovignetten aus schulischem Musikunterricht.

Um die Wirksamkeit des Kurskonzepts zu überprüfen, wurde ein Vignettentest entwickelt und im Rahmen eines quasi-experimentellen Designs (Prä-Post-Studie mit Versuchs- und Kontrollgruppe) insgesamt 55 Studierenden vorgelegt. Die Ergebnisse legen nahe, dass sich mit Hilfe der fallbasierten Arbeit im Kurs tatsächlich ein Lernzuwachs in den angestrebten Bereichen erzielen lässt. Zudem könnte der Vignettentest über Evaluationszwecke hinaus auch für Grundlagenforschung im Bereich professioneller Kompetenzen nutzbar sein.

How could future music teachers be adequately prepared for singing with heterogeneous groups in the general music classroom? In which way is it possible to monitor future music teachers' progress? As part of the "Qualitätsoffensive Lehrerbildung" funded by the German Ministry of Education and Research, the Chair of Music Education at the University of Augsburg is conducting a development and research project that aims to foster professional knowledge and situation-specific skills of student teachers. An essential element of the newly developed seminar is case-based learning, using audio and video vignettes from music lessons. In order to empirically investigate the effectiveness of this concept, we administered a vignette-based assessment tool to a sample of 55 music teacher students in a pre-post-study with treatment and control groups. The results indicate that the students participating in the course achieved substantial learning gains in the target-

ed competencies. Furthermore, the test concept has proven suitable for evaluation purposes and might also be fruitful for basic research.

Einleitung

Im Rahmen der „Qualitätsoffensive Lehrerbildung“ wird an der Universität Augsburg das Projekt „Förderung der Lehrerprofessionalität im Umgang mit Heterogenität“ (LeHet) durchgeführt. Das Teilprojekt „LeHet Musik“ befasst sich mit Entwicklung und Implementierung eines Lehrformats für den Bereich „Vokaldidaktik/Singen mit heterogenen Gruppen im Klassenunterricht Musik“. Neben domänenspezifischem Professionswissen sollen situationsspezifische Fähigkeiten der Studierenden und situationsgerechter Einsatz künstlerisch-musikpraktischer Fertigkeiten angebahnt und gefördert werden. Ein wesentliches methodisches Element des Kurskonzepts ist fallbasiertes Arbeiten mit Video- und Audiovignetten aus schulischem Musikunterricht.¹

Zur empirischen Überprüfung der Wirksamkeit des neuen Lehrformats wurde ein Testinstrument zur Messung von Professionswissen und situationsspezifischen Fähigkeiten entwickelt, das an Vorarbeiten aus dem Projekt FALKO-M anknüpft (Puffer & Hofmann, 2017). In dessen Rahmen war ein fachspezifischer Professionswissenstest konzipiert worden, der unter anderem Unterschiede zwischen Expert*innen und Noviz*innen sowie zwischen Studierenden und Lehrkräften verschiedener Ausbildungsrichtungen messbar machte. Im Zuge der Konstruktion des LeHet-Musik-Tests stellte sich unter anderem die Frage, inwieweit sich ein Testformat, das zur Erforschung vergleichsweise ‚großräumiger‘ Unterschiede in den professionellen Kompetenzen von Musiklehrkräften entwickelt wurde, auch zur Evaluation von Lernzuwächsen bei Studierenden während eines fachdidaktischen Seminars als geeignet erweisen würde, also für die Messung von Veränderungen über einen Zeitraum von wenigen Wochen hinweg.

Theoretischer Rahmen

Anknüpfend an kognitionspsychologisch grundierte Konzepte wird (Musik-) Unterricht im Projekt LeHet Musik als Sequenz von Handlungssituationen aufgefasst, mit denen sich Lehrkräfte aktiv auseinandersetzen. Dabei lösen sie eine

1 Das diesem Bericht zugrundeliegende Vorhaben wurde im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1509 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin. Details zum Projekt unter <https://www.uni-augsburg.de/de/fakultaet/philsoz/fakultat/musikpaedagogik/forschung/lehet-musik/>

Folge von Handlungsproblemen (vgl. Bromme, 1992, 1997; Dann & Haag, 2017). Die dabei notwendigen Aktivitäten beinhalten beispielsweise jene fortwährenden größeren und kleineren Planungsanpassungen, die zu den Erfordernissen adaptiven Unterrichtens gehören (Göllner & Niessen, 2016). Zur erfolgreichen Bewältigung solcher Herausforderungen benötigen Musiklehrkräfte ein umfangreiches Bündel professioneller Kompetenzen. Den konzeptionellen Gesamtrahmen für Lehrkonzept und Evaluation in LeHet Musik bildet ein überfachliches Modell professioneller Kompetenz von Lehrkräften nach Blömeke, Gustafsson und Shavelson (2015), in dem die dispositionsbezogenen Komponenten bisheriger Modelle um situationsspezifische Fähigkeiten erweitert sind:

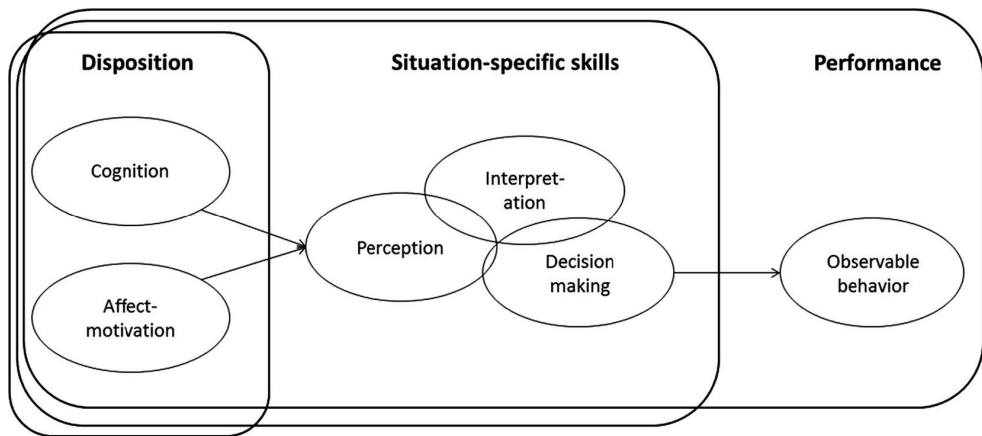


Abbildung 1: Professionelle Kompetenz von Lehrkräften als Kontinuum (Blömeke et al., 2015, S. 7)

LeHet Musik nimmt das Professionswissen angehender Musiklehrkräfte als eine wesentliche Grundlage erfolgreichen Unterrichtens in den Blick (vgl. Baumert & Kunter, 2011; Blömeke et al., 2011; Mulder & Gruber, 2011; Krauss, Lindl & Schilcher, 2017). Handlungsnahe professionelle Wissensbestände, die für das Singen mit Kindern im Klassenverband notwendig erscheinen, sind unter Rückgriff auf das FALKO-M-Modell zu Fachwissen und Fachdidaktischem Wissen von Musiklehrkräften konzeptualisiert (Puffer & Hofmann, 2017).

Handreichungen zum Singen mit Kindern (z. B. Mohr, 2013; Wieblitz, 2013) setzen in der Regel besondere situationsspezifische Fähigkeiten voraus, beispielsweise diagnostisches Sehen und Hören. Auch diese sollen im LeHet-Musik-Seminar angebahnt und gefördert werden. Als Grundlage einer entsprechenden Modellierung wird das Konzept professioneller Unterrichtswahrnehmung (*professional vision*) herangezogen (Sherin, 2007; Sherin & van Es, 2009; Seidel, Blomberg & Stürmer, 2010; Hellermann, Gold & Holodyski, 2015). Diese lässt sich im Rahmen eines ‚Als-Ob-Modells‘ in zwei Teilprozesse untergliedern: Wahrnehmung im engeren Sinne (*selective attention/noticing*) und Interpretation (*know-*

ledge-based reasoning). Wissensgesteuerte Aufmerksamkeit (*selective attention*) ermöglicht die Identifikation (*noticing*) solcher Situationen und Ereignisse im Unterricht, die aus professionellem Blickwinkel relevant sein könnten für den Erfolg von Unterrichtshandlungen. Die Komponente *knowledge-based reasoning* bezeichnet die Fähigkeit, wahrgenommene Ereignisse auf der Basis vorhandener Wissensbestände zu analysieren und zu interpretieren. Dies ermöglicht eine diagnostische Einschätzung der Situation, auf deren Grundlage dann didaktisch angemessene² Handlungsentscheidungen getroffen werden können.³

Im hochschuldidaktischen Kontext gilt es weiterhin zu unterscheiden zwischen professioneller Wahrnehmung aus der ‚Innensicht‘, also der Perspektive der Lehrkraft, die unter unmittelbarem Handlungsdruck beobachtet, analysiert und Entscheidungen trifft, und Wahrnehmungen aus der ‚Außensicht‘ von Unbeteiligten, die Unterricht beobachten (live oder medial vermittelt) und sich ohne Handlungsdruck der Analyse des Geschehens und dem Entwerfen von Handlungsmöglichkeiten und -alternativen widmen können (Hellermann et al., 2015, S. 98). Grundlegend für das hochschuldidaktische Konzept von LeHet Musik ist die Annahme, dass sich die in der Außensicht durch Auseinandersetzung mit Video- und Audiobeispielen gewonnenen Erkenntnisse und die dabei trainierten situationsspezifischen Fähigkeiten zumindest in gewissem Umfang für das spätere eigene Unterrichtshandeln nutzen lassen (Konzept der ‚Annäherungen‘ an professionelle pädagogische Praxis; vgl. Grossman et al., 2009; Heinzl & Krausemann, 2014; Hellermann et al., 2015; Seidel & Thiel, 2017; kritisch z.B. Dann & Haag, 2017, S. 114).

Konzeptualisierung der domänenspezifischen Kompetenzen

Im Fokus von Kurskonzept und Evaluation des Projekts LeHet Musik stehen Kompetenzen zur Bewältigung von Anforderungen, wie sie sich Musiklehrkräften beim Singen mit Klassen in den Jahrgangsstufen 1 bis 6 stellen. Als heuristische Basis zur inhaltlichen Konkretisierung der notwendigen Wissensfacetten und situationsspezifischen Fähigkeiten dienen schulische Curricula (z.B. Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2014), die Ländergemeinsamen inhaltlichen Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung der deutschen Kultusministerkonferenz (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik

2 Als ‚didaktisch angemessen‘ wird eine Unterrichtshandlung in der vorliegenden Studie aufgefasst, wenn sie entsprechenden Kriterien ‚guten Unterrichts‘ aus musikdidaktischer Literatur entspricht (zum Vorgehen: vgl. Krauss et al., 2011, S. 144; zum Konzept ‚guten Unterrichts‘: siehe Berliner, 2005).

3 Eine Zusammenfassung verschiedener kognitionspsychologischer Modelle von Lehrerhandeln bieten z.B. Dann & Haag, 2017, S. 103–107.

Deutschland [KMK], 2019, S. 41–42) und musikdidaktische Literatur zum Singen mit Kindern (z.B. Mohr, 2013; Arnold-Joppich, 2015; Fuchs, 2015; Henning, 2014; Hofmann, 2015; Wieblitz, 2013).

Aus dem Bereich der kognitiven Dispositionen (vgl. Abb. 1) werden drei Facetten fachdidaktischen Wissens in den Blick genommen: Wissen über typische Lernwege und -schwierigkeiten von Kindern beim Singen und über Möglichkeiten entsprechender Diagnostik, Wissen über geeignete Instruktionsstrategien sowie Kenntnisse über Klassenführung und Unterrichtsorganisation in Musiziersituationen (vgl. dazu auch Puffer & Hofmann, 2017, S. 253–255). Hinzu kommt Fachwissen zu Stimmphysiologie und zu stimmlicher Entwicklung von Kindern und Jugendlichen.

Erforderliche situationsspezifische Fähigkeiten umfassen das Vermögen, die Aufmerksamkeit gezielt auf Unterrichtsergebnisse und konstellationen zu richten, die für den Lernerfolg einzelner Schüler*innen oder der gesamten Lerngruppe sowie für den weiteren Fortgang des Unterrichtsgeschehens von Bedeutung sein könnten (*selective attention*). Wissensgeleitet wahrgenommen werden sollen einerseits Verhaltensweisen der Kinder, die beides positiv wie negativ beeinflussen können, andererseits stimmlich-musikalische Qualitäten in den Dimensionen Körperspannung/Haltung/Bewegung, Atemführung und Phrasierung, Phonation/Artikulation/Stimmklang, Intonation, Modulation und Ausdruck, Rhythmus und Tempo (*noticing*). Bei der Unterrichtswahrnehmung aus der ‚Außensicht‘ im Rahmen der Fallarbeit im Seminar tritt die Fähigkeit hinzu, das Beobachtete präzise und differenziert zu beschreiben.

Die Kompetenzfacette ‚Interpretation‘/*knowledge based reasoning* wird in zwei Teilkomponenten modelliert, die sich in Beziehung zu den eben skizzierten Komponenten fachdidaktischen Wissens setzen lassen: einen diagnostischen Anteil und einen, der sich auf das professionelle Abwägen von Handlungsalternativen richtet. Im ersten Schritt geht es darum, die wahrgenommenen Ereignisse auf der Grundlage der professionellen Wissensbasis zu erklären, zu deuten oder zu beurteilen – beispielsweise auf Basis der Kenntnisse typischer Lernschwierigkeiten von Kindern beim Singen in der Großgruppe. Im zweiten Schritt werden, wiederum wissensbasiert, Möglichkeiten professionellen Handelns bzw. Reagierens entwickelt und gegeneinander abgewogen. Auf dieser Basis können schließlich didaktisch angemessene, sach- und schüler*innengerechte Handlungsentscheidungen getroffen werden, beispielsweise für das Beginnen, Fortsetzen, Modifizieren oder den Abbruch einer Übung, für Planungsanpassungen im Rahmen des aktuellen Unterrichtsgeschehens, oder für längerfristig durchzuführende Maßnahmen im Rahmen der weiteren Unterrichtsplanung und -vorbereitung für die Lerngruppe (z.B. das regelmäßige Durchführen von Übungen zum Erschließen der Kopfstimme).

Ziele der Evaluationsstudie

Im Rahmen einer Evaluationsstudie zu LeHet Musik galt es zu überprüfen, ob Teilnehmer*innen des ‚LeHet-Seminars‘ am Ende der Vorlesungszeit tatsächlich über eine umfangreichere domänenspezifische Wissensbasis und über besser entwickelte situationsspezifische Fähigkeiten verfügen als zu Beginn sowie über eine bessere Fähigkeit, beides in didaktisch angemessene Handlungsentscheidungen umzusetzen. Hypothese (1) lautet deshalb, dass Studierende nach der Teilnahme am Seminar in einem entsprechenden Leistungstest signifikant und mit mindestens mittlerer Effektstärke höhere Testscores erzielen als zu Beginn. Zudem interessiert die Frage, ob eventuelle Leistungszuwächse tatsächlich durch die Teilnahme am Seminar beeinflusst sind. Hypothese (2) nimmt an, dass sich bei Studierenden einer Vergleichsgruppe kein oder nur ein deutlich geringerer entsprechender Zuwachs in den interessierenden Kompetenzfacetten messen lässt.

Ein weiteres Forschungsanliegen entstand im Zuge der Arbeit am Forschungsprojekt FALKO-M. Der Großteil der dort verwendeten Testaufgaben zum fachdidaktischen Wissen von Musiklehrkräften und Lehramtsstudierenden ist in offenem Itemformat gehalten (Puffer & Hofmann, 2017, S. 256–257). Bei der Auswertung dieses Aufgabentyps fiel auf, dass Studierende im Vergleich zu den teilnehmenden Lehrkräften deutlich geringere Testscores erzielten, obwohl ihre Testantworten ähnlich umfangreich ausfielen. Häufig eignete sich nur ein verhältnismäßig kleiner Anteil der Testantworten der Studierenden tatsächlich zur Bewältigung der jeweils präsentierten Aufgabe. Die übrigen Antworten entstammten zwar meist der passenden Wissenskategorie (z. B. Wissen über typische Schüler*innenfehler beim Singen), passten aber nicht zum konkreten Problem. Dies lässt sich als Indikator dafür interpretieren, dass die teilnehmenden Studierenden bereits durchaus über eine Basis an fachspezifischem Professionswissen verfügten, das Wissen aber nur in vergleichsweise geringem Umfang situations- und problemgerecht einsetzen konnten. Auch diese Fähigkeit sollte sich bei den Teilnehmer*innen des LeHet-Seminars verbessern. Deshalb wird zusätzlich angenommen (Hypothese 3), dass die Studierenden des LeHet-Seminars bei der Testwiederholung einen höheren Anteil korrekter Antworten an den insgesamt gegebenen Testantworten erreichen als bei der ersten Testung.

Methodisches Vorgehen

Die Überprüfung der Wirksamkeit des Seminarkonzepts erfolgte im Sommersemester 2018 in Form eines quasi-experimentellen Vorgehens (Prä-Post-Design mit Versuchs- und Kontrollgruppe). Unabhängige Variable ist damit die Zugehörigkeit zu einer der beiden Gruppen. Die Versuchsgruppe bestand aus den

Studierenden des LeHet-Seminars. Als Kontrollgruppe standen die Teilnehmerinnen eines parallel stattfindenden vokaldidaktischen Seminars mit Schwerpunkt auf dem Erlernen eines Solmisationsverfahrens zur Verfügung. Aus organisatorischen Gründen wurde die Vergleichsgruppe von einem anderen Dozenten unterrichtet als die Versuchsgruppe. Beide Seminare erhielten wöchentlich 90 Minuten Präsenzunterricht; zusätzlich bearbeiteten die Studierenden jeweils häusliche Vor- und Nachbereitungsaufgaben (beide Gruppen: Lektüreaufgaben, Vorbereiten von Einstudierungen mit der Seminargruppe; Versuchsgruppe: geringere Gewichtung dieser Aktivitäten zugunsten des individuellen Bearbeitens von Beobachtungs- und Diagnoseaufgaben an Unterrichtsvideos mit Hilfe einer Online-Plattform; Kontrollgruppe: keine Verwendung von Video- oder Audioaufnahmen aus schulischem Musikunterricht). Zu Beginn und gegen Ende der Kurse füllten die Teilnehmer*innen je einen Leistungstest zu domänenspezifischem Professionswissen und situationsspezifischen Fähigkeiten aus. Abhängige Variablen der Studie sind somit die erreichten Testscores in Prä- und Posttest sowie eine Kenngröße, die für jede*n Teilnehmer*in den erzielten Anteil korrekter bzw. didaktisch angemessener Antworten in Relation zu den insgesamt von ihr oder ihm gegebenen Antworten markiert (siehe unten).

Stichprobe

Es konnte eine Vollerhebung durchgeführt werden. Insgesamt beteiligten sich 55 Studierende an den Tests (Erhebung 1: $n = 50$, Erhebung 2: $n = 32$). In die Testauswertung gingen die Daten derjenigen Studierenden ein, die an beiden Erhebungen teilnahmen ($N = 29$; Versuchsgruppe $n = 18$, Kontrollgruppe: $n = 11$). Aus der Versuchsgruppe stehen damit die Daten von 3 Männern und 15 Frauen zur Verfügung, von denen 12 Lehramt GS studierten (66,7%), 5 Lehramt Mittelschule (27,8%) und eine*r Lehramt Realschule; 4 Teilnehmer*innen studierten Musik als Unterrichtsfach, 14 (77,8%) hatten Musik als Didaktikfach belegt.⁴ Aus der Kontrollgruppe beteiligten sich 11 Frauen an beiden Tests. Auch hier bildeten Studierende des Lehramts Grundschule sowie Studierende mit Musik als Didaktikfach eine deutliche Mehrheit (Lehramt GS: 8 (72,7%), MS = 0, RS = 2 (18,2%), k.A. = 1; Unterrichtsfach Musik: 4 (36,4%), Didaktikfach Musik: 7 (63,6%)).

Die Studierenden waren durchschnittlich 21,03 Jahre alt ($SD = 2,56$) und studierten im dritten Fachsemester ($M = 3,14$, $SD = 1,65$). Der Durchschnitt ihrer Abiturnoten, die als grobe Indikatoren für die allgemeine kognitive Leistungs-

4 Das Studium von Musik als ‚Dritteldidaktikfach‘ ist in Bayern für das Lehramt an Grund- oder Mittelschulen möglich. Es kann ohne Eignungsprüfung begonnen werden. An der Universität Augsburg sind dafür 12 (Lehramt GS) bzw. 20 (Lehramt MS) ECTS-Punkte vorgesehen (Universität Augsburg, 2013, S. 112; Übersicht zur Struktur der Lehramtsstudiengänge in Bayern: Krauss, Lindl, Schilcher & Tepner, 2017, S. 52–54).

fähigkeit herangezogen werden können (Baron-Boldt, Schuler & Funke, 1988; Kunina-Habenicht et al., 2013), lag bei 2.23 ($SD = .38$) und damit etwas über dem Wert, der in anderen Untersuchungen für Studierende nicht vertiefter Lehrämter ermittelt wurde (z. B. Klusmann, 2011; Kunina-Habenicht et al., 2013).

Durchführung

In der zweiten Woche der Vorlesungszeit und ein weiteres Mal nach 10 Seminarsitzungen wurde den Studierenden ein vignettenbasierter Papier- und Bleistift-Test zu Professionswissen und situationsspezifischen Fähigkeiten mit einer Bearbeitungszeit von etwa 30 Minuten vorgelegt. Diese Testdauer erwies sich als vertretbarster Kompromiss zwischen dem Erkenntnisinteresse und den Erfordernissen des Lehrbetriebs: Die Tests fanden während der Seminarsitzungen statt, die damit einhergehende zeitliche Belastung sollte nicht über Gebühr hoch ausfallen. Je eine Projektmitarbeiterin führte den Test am Beginn der jeweiligen Seminarsitzung durch. Die im Test verwendeten Audio- und Videovignetten sowie die Art der Aufgabenstellung waren für alle Teilnehmer*innen neu und nicht zuvor in einem der Seminare verwendet worden. Die Video- und Audio-Stimuli wurden jeweils zweimal hintereinander per Projektion und Stereoanlage präsentiert. Die jeweilige Testleiterin achtete darauf, die Bearbeitung einer Aufgabe erst zu beginnen, wenn alle Teilnehmer*innen die vorhergehende Aufgabe abgeschlossen hatten; außerdem protokollierte sie den Ablauf jeder Testsitzung schriftlich. Die Tests wurden anonym durchgeführt. Um dennoch eine Zuordnung der Testleistungen aus Prä- und Post-Test zu ermöglichen, wurde für jede*n Teilnehmer*in ein Pseudonym erstellt: ein achtstelliges Codewort aus Daten, die für die jeweilige Person jederzeit replizierbar sind, zu denen das Forschungsteam jedoch keinen Zugang hat.

Erhebungsinstrument

Testkonzept

Tabelle 1 enthält einen Überblick über die Items des LeHet-Musik-Tests. Domänenspezifisches Fachwissen wird mit Hilfe eines Multiple-Choice-Items gemessen, die skizzierten Facetten fachdidaktischen Wissens und situationsspezifischer Fähigkeiten werden mit Hilfe vignetten-basierter Items mit offenem Antwortformat erfasst.

Tabelle 1: Vignettentest zur Erfassung von Professionswissen und situationsspezifischen Fähigkeiten: Übersicht über Verteilung der Items, Stimuli und Antwortformate

Aufgabenstamm	Items	Stimulus	Antwortformat
„Guten Morgen“	1,2	Audio	offen
„Brummer“	3	Audio	geschlossen (Quasi-Paarvergleich) ⁵
„Kookaburra“	4,5	Audio	offen
„O Happy Day“	6,7	Video	offen
„Dreiklangsbrechungen“	8,9	Video	offen
„Singen und Stimme“	10		geschlossen (Multiple Choice)

Bem.: SK: Schülerkognitionen, Lernwege & Lernschwierigkeiten; MI: Musikbezogene Instruktionsstrategien; FW: Fachwissen

Offene Antworten erfordern das aktive Konstruieren von Bedeutungszusammenhängen und erlauben Rückschlüsse auf die Beschaffenheit der dahinter liegenden Wissensgrundlagen, Wahrnehmungs- und Denkprozesse (Hamann & Jördens, 2014, S. 169–170; Mayer & Wellnitz, 2014, S. 27). Das Testkonzept basiert auf der Annahme, dass sich Unterschiede in den interessierenden kognitiven Leistungen in Umfang, Differenziertheit und Korrektheit bzw. didaktischer Angemessenheit der Testantworten niederschlagen.

Abbildung 2 zeigt beispielhaft eine der offenen Testaufgaben:⁶

„Oh Happy Day“ Im Videobeispiel singt eine Schulklasse das Lied „Oh Happy Day“.	
Welche Probleme beim Singen können Sie bei den Kindern erkennen? ⁷	Wie würden Sie diesen Problemen in Ihrem Unterricht abhelfen?

Abbildung 2: Beispielimitem „Oh Happy Day“

5 Zu Aufgabenformat und Auswertungsverfahren: siehe Puffer & Hofmann, 2017, S. 264.

6 Das vollständige Testheft ist abrufbar unter <https://www.uni-augsburg.de/de/fakultaet/philoz/fakultat/musikpaedagogik/forschung/lehret-evaluation/>

7 Die beiden Items wurden gemeinsam mit Studierenden entwickelt. In der Pilotierung zeigten die ersten Testteilnehmer*innen starke Tendenzen, die Aufmerksamkeit beim Betrachten der Videovignette auf die Lehrkraft zu richten anstatt auf die Schüler*innen. Die vergleichsweise starke ‚Lenkung‘ in Aufgabenformulierung und -formatierung erwies sich als geeigneter Weg, dem entgegenzuwirken.

Auswertung

Die offenen Items werden nach einem quantitativ-inhaltsanalytischen Verfahren von je zwei geschulten Rater*innen kodiert. Grundlage dieser Arbeit ist ein detailliertes Kodiermanual mit insgesamt 10 Antwortkategorien, die den Testitems je nach Aufgabenstellung zugewiesen werden (siehe Abb. 3). 7 Kategorien betreffen stimmbildnerische Aspekte (Körperspannung, Haltung, Bewegung; Atemführung, Phrasierung; Phonation & Intonation; Artikulation; Stimmklang, klangliche Qualitäten; Rhythmus & Tempo; Modulation & Ausdruck). Damit soll der fundamentalen Bedeutung Rechnung getragen werden, die entsprechenden diagnostischen Fähigkeiten von Lehrkräften beim Singen mit Kindern zukommt (vgl. z. B. Mohr, 2013, S. 35; Maschke, 2008, S. 15–17). Die 3 verbleibenden Kategorien beziehen sich auf fachdidaktisches Wissen über domänenspezifische Aspekte von Klassenführung und Unterrichtsorganisation, die angemessene Berücksichtigung von Kontexten, denen ein Lied entstammt oder in die es eingebettet werden kann (vgl. Hofmann, 2015, S. 106–108; Fuchs, 2015, S. 127–128) sowie die Fähigkeit, mögliche pädagogische Anschlusshandlungen im Sinne eines systematisch aufbauenden methodischen Vorgehens strukturiert und kohärent darzustellen (vgl. dazu das Vorgehen in Puffer & Hofmann, 2017, S. 262). Die Ermittlung des jeweiligen Testscores (erste Erhebung: abhängige Variable 1, zweite Erhebung: abhängige Variable 2) erfolgt über das Zählen und Aufsummieren der als diagnostisch korrekt bzw. didaktisch angemessen bewerteten, substanziell unterschiedlichen Antworten (vgl. Abb. 3). Dabei wird keine qualitative Gewichtung der einzelnen Kategorien gegeneinander vorgenommen.

Aufgabenstellung: Welche Probleme beim Singen können Sie bei den Kindern erkennen?			
Inhaltliche Leistung	Beispiel	Ant- wort- code	Score
Keine Antwort	Gar nichts, Strich oder „?“	999	0
Falsche/unverständliche/unvollständige Lösung bzw. Aufgabe missverstanden	Frage missverstanden: Teilnehmer*in zählt z. B. Fehler auf, die nichts mit Singen zu tun haben (z. B. „Xylophone spielen nicht in time“, „Lehrer kann die Rhythmusgruppe nicht sehen“); Mittel zur Abhilfe werden genannt (die in die Lösung zum zweiten Item gehören), z. B. „Stimmbildungsübungen für kräftigeren Sound“ Falsche Lösung: z. B. „Mädchen singen zu laut“ Unvollständige Lösung: z. B. nur Schlagworte wie „Technik“, „Intonation“	0	0
„Power-Aufgabe“ mit nach oben offener Höchstpunktzahl!			
Gesamtzahl Antworten	Gesamtzahl aller inhaltlich voneinander verschiedenen ‚diagnostischen‘ Antworten (ohne Rücksicht darauf, ob sie korrekt sind)	777	Score
Richtige Lösungen: Kategorien	(Fachbegriffe müssen nicht in dieser Form genannt werden, Umschreibung genügt)		
Körperspannung, Haltung, Bewegung	– Insgesamt hypotone Singhaltung (zu wenig Körperspannung, Blickrichtung z. T. nach unten hängende Schultern etc.); bzw.: Körperspannung wird nicht aufrechterhalten – Bewegungsaktivitäten wirken ungezielt, stören eher als Stimmtechnik und Ausdruck zu fördern	1	Score = Anzahl der unter Antwortcode 1-8 richtig benannten, inhaltlich voneinander verschiedenen Einzelaspekte. Pro Code können auch mehrere Punkte vergeben werden!
Intonation	– eingetrübte Intonation (als Folge mangelnder Körperspannung)	2	
Artikulation	– Text stellenweise nicht gut verständlich, undeutliche, verwaschene Aussprache (insbes. Konsonanten) – Mehrfach „Zwischen-h“ statt Legato (z. B. „Day-hay“, „wha-hashed“)	3	
[...]			

Abbildung 3: Auszug aus der Kodieranweisung zum ersten Item aus der Beispielaufgabe „O Happy Day“ (siehe Abb. 2). Es folgen noch vergleichbar detaillierte Anweisungen zu den Kategorien Stimmklang, klangliche Qualitäten (4), Modulation & Ausdruck (5), Rhythmus & Tempo (6), Kontext (7), Organisatorisches (8).

Um Aufschluss über den Anteil korrekter bzw. didaktisch angemessener Antworten an den insgesamt gegebenen Testantworten zu erhalten, werden für jedes offene Item die insgesamt gegebenen, substantiell verschiedenen und kategorial passenden („diagnostischen“ oder „methodischen“) Antworten gezählt – unabhängig davon, ob sie im Sinne des Kodiermanuals korrekt sind oder nicht (Antwortcode 777). Anschließend wird ein Quotient berechnet aus dem erreichten Testscore und der Zahl der Testantworten, die der Teilnehmer bzw. die Teilnehmerin insgesamt zum jeweiligen Item gegeben hat (Kenngröße für den Anteil korrekter Testantworten, abhängige Variable 3). Notiert also beispielsweise eine Studentin beim Bearbeiten des ersten Items der Aufgabe „Oh Happy Day“ drei korrekte Beobachtungen zu Problemen der Kinder beim Singen und außerdem noch „Die Mädchen singen zu laut“, also eine falsche Antwort, ergibt sich ein Quotient von 0.75 bzw. eine ‚Trefferquote‘ von 75% (3 korrekte aus 4 insgesamt gegebenen Antworten). Der Wert dieser Kenngröße kann zwischen 0 (keine korrekte Antwort) und 1 (alle gegebenen Antworten sind korrekt) liegen.

Psychometrische Kennwerte

Tabelle 2: Anhaltspunkte zur psychometrischen Qualität des Testinstruments

Skala: Professionelle Wahrnehmung und Anwendung fachdidaktischen Wissens (Items mit Vignetten + offenen Antwortformaten, $n = 8$)			Gesamttest
Auswertungsobjektivität (Interraterreliabilität)	Skalenreliabilität	Trennschärfe	Augenscheinvalidität (Ausbildungsrelevanz)*
82 doppelt kodierte Testhefte von 55 Studierenden (Erhebung 1 und 2)	$N = 48$ (32 Studierende, 2. Erhebung und 16 Lehrkräfte aus der Testvalidierung)		$n = 16$ (Lehrkräfte aus der Testvalidierung)
Spearman's ρ : M (SD)	Cronbachs α	r_{it} : M (SD)	M (SD)
.88 (.08) range [.71–.96]	.92	.76 (.06) range [.67–.84]	3.67 (.14) range [3.44–3.81]

* Skalierung Augenscheinvalidität (Relevanz der Szenarien und Aufgaben für die Ausbildung von Musiklehrkräften): 1 Nicht relevant, 2 Kaum relevant, 3 Ziemlich relevant, 4 Sehr relevant

Tabelle 2 fasst einige psychometrische Kennwerte des Testinstruments zusammen. Aus den Werten lässt sich der Schluss ziehen, dass sich der Test objektiv auswerten lässt (Interraterreliabilität)⁸ und dass alle Items etwas Ähnliches

8 Die Wahl von Spearman's ρ als Kenngröße erfolgte unter der Annahme, dass die ermittelten Scores Ordinalniveau besitzen (siehe Tab. 2; vgl. Wirtz & Caspar, 2002, S. 133–135).

messen (Skalenreliabilität und Trennschärfen). Zudem wurden die Testaufgaben nicht nur von Expert*innen in der Pilotierung, sondern auch von 16 zu Validierungszwecken getesteten berufserfahrenen Musiklehrkräften als inhaltlich valide beurteilt (Inhalts- und Augenscheinvalidität).

Statistisch abgesicherte Aussagen über die Testgüte sind auf Grundlage der geringen Anzahl der Testteilnehmer*innen nicht möglich. Die im Folgenden berichteten Befunde sind mit entsprechender Vorsicht zu interpretieren.

Datenaufbereitung

Zunächst wurden alle Testhefte wie oben beschrieben kodiert und die Rater-Übereinstimmung mit Hilfe von Spearmans ρ kontrolliert. Anschließend wurden auf Basis der Teilnehmerpseudonyme alle Testhefte aus dem Datensatz entfernt, deren Bearbeiter*innen nur an einer der beiden Erhebungen teilgenommen hatten. Es folgte eine Überprüfung auf fehlende Werte. In die weiteren Analysen gingen $2 \times 29 = 58$ vollständig ausgefüllte Testhefte ein. Für jede Teilnehmerin bzw. jeden Teilnehmer wurde durch Aufsummieren aller in den Einzelitems erreichten Scores je ein Gesamtscore für den Prä- und Post-Test berechnet (Werte der abhängigen Variablen 1 und 2). Eine Gewichtung der Items gegeneinander wurde dabei nicht vorgenommen. Eine Überprüfung auf Ausreißer nach oben oder unten ergab keinen Befund. Der Wert für die abhängige Variable 3 („Trefferquote“) wurde ermittelt, indem über alle 8 offenen Items hinweg der Mittelwert der Kenngröße für den Anteil korrekter Antworten berechnet wurde. Außerdem wurde durch Zuweisung eines entsprechenden Werts für die unabhängige Variable kenntlich gemacht, welche Testhefte der Versuchs- (1) und welche der Kontrollgruppe (0) entstammten. Die weitere Auswertung der so aufbereiteten Daten erfolgte mit Hilfe von SPSS.

Datenanalyse

Zunächst wurde sichergestellt, dass die Verteilungen der Testwerte in beiden Erhebungen nicht signifikant von einer Normalverteilung abweichen (Überprüfung mittels Histogrammen, Q-Q-Plots und Shapiro-Wilks-Test; $p > .05$). Mit Hilfe von t-Tests für unabhängige Stichproben wurde zudem überprüft, ob sich Versuchs- und Kontrollgruppe hinsichtlich ihres Alters, des Fachsemesters oder der Abiturnote unterscheiden. Es ergab sich für keines der drei Merkmale ein signifikanter Unterschied.

Die Hypothesen 1 und 2 wurden jeweils mit Hilfe von t-Tests für abhängige Stichproben getestet (Zugehörigkeit zu Versuchs- bzw. Kontrollgruppe als unabhängige Variable, Testscores der Erhebungen 1 und 2 als abhängige Variablen).

Um zu klären, ob sich die Testscores vom Prä- zum Post-Test in den beiden Gruppen statistisch signifikant unterschiedlich veränderten, wurde eine Varianzanalyse mit Messwiederholung durchgeführt (Zugehörigkeit zu Versuchs- bzw. Kontrollgruppe als Zwischensubjektfaktor, Testscores der Erhebungen 1 und 2 als Innersubjektfaktoren).

Zur Prüfung von Hypothese 3 kamen ebenfalls t-Tests für abhängige Stichproben zum Einsatz (Zugehörigkeit zu Versuchs- bzw. Kontrollgruppe als unabhängige Variable; Kenngrößen für den Anteil korrekter Antworten aus den Erhebungen 1 und 2 als abhängige Variable).

Um zudem die Testergebnisse der Studierenden im Kontext des Professionalisierungsprozesses grob einordnen zu können, wurden zwei explorative Post-hoc-Analysen vorgenommen: Die Leistungen der Studierenden (Testscores und Anteil der korrekten Antworten) wurden mit jenen der berufserfahrenen Musiklehrkräfte verglichen, die den Test zu Validierungszwecken ausgefüllt hatten.

Ergebnisse

Mit Hilfe des Testverfahrens sollte die Frage beantwortet werden, ob sich im Laufe eines mehrwöchigen Seminars Wissensbestände, professionelle Wahrnehmung sowie die Fähigkeit, beides im Rahmen video- oder audiobasierter Fallarbeit in korrekte Diagnosen und didaktisch angemessene Handlungsentscheidungen umzusetzen, messbar verbessern lassen. Ein Vergleich der Gesamtscores der Testleistungen von Versuchs- und Kontrollgruppe stützt die Annahme, dass dies tatsächlich der Fall ist. Beide Gruppen konnten in der zweiten Messung deskriptiv bessere Leistungen erzielen als in der ersten. Bei den Mitgliedern der Versuchsgruppe fiel der Unterschied groß aus: Sie erreichten in der ersten Erhebung im Durchschnitt 13,58 Punkte ($SD = 3.07$); in der zweiten Erhebung lag das durchschnittliche Ergebnis bei 16,50 Punkten ($SD = 4.30$; $t(17) = 2.81$, $p < .01$ (einseitig), $d_{\text{repeated measures}} = 0.82$, 95% KI [0.135, 1.495]).⁹ Dies kann als Indikator für die Gültigkeit von Hypothese 1 gelten. Bei der Kontrollgruppe ergab sich dagegen kein signifikanter Unterschied zwischen den beiden Erhebungen, was konform mit Hypothese 2 ist (erste Erhebung: $M = 14.64$, $SD = 3.43$; zweite Erhebung: $M = 15.41$, $SD = 4.16$; $t(10) = 0.49$, $p = .64$ (zweiseitig), $d_{\text{repeated measures}} = .16$; 95% KI [-0.674, 1]). Auch das Ergebnis der Varianzanalyse fiel signifikant aus, mit mittlerer bis großer Effektstärke ($F(1, 27) = 4.13$, $p = .05$, partielles $\eta^2 = .13$). Angesichts der kleinen Datengrundlage ist dieser Befund allerdings nur bedingt aussagekräftig.

Der rechnerisch große Effekt beim Vergleich der Testleistungen der Versuchsgruppe sollte indes nicht darüber hinwegtäuschen, dass das Ausgangs-

⁹ Berechnung der Effektstärke bei wiederholten Messungen hier wie im Folgenden nach Morris und DeShon (2002).

niveau aller Studierenden bezüglich der zu testenden Fähigkeiten ein niedriges war. Nimmt man den gesamten Professionalisierungsprozess in den Blick, so führt auch die von der Versuchsgruppe erzielte Steigerung nur zu Leistungen auf einem immer noch recht niedrigen Niveau. Dies lässt sich aus einer der vorgenommenen Post-hoc-Analysen folgern: Vergleicht man die durchschnittlichen Testergebnisse aller Studierenden aus der zweiten Erhebung mit jenen der berufserfahrenen Lehrkräfte, die den LeHet-Musik-Test zu Validierungszwecken ebenfalls ausfüllten, so zeigen sich enorme Unterschiede zwischen den Leistungen der Noviz*innen ($n = 29$, $M = 16.09$, $SD = 4.21$) und denen der Expert*innen ($n = 16$, $M = 42.97$, $SD = 13.95$; $t(16.52) = 7.52$, $p < .001$ (einseitig), $d = 3.02$; 95% KI [2.144, 3.888]; siehe Abb. 4).

Geht man von der Annahme aus, dass berufserfahrene Lehrkräfte über ein größeres diagnostisches und didaktisches Repertoire sowie über besser trainierte und differenzierte professionelle Wahrnehmungsfähigkeiten verfügen als Studierende, so entspricht dieser Unterschied den Erwartungen und lässt sich im Sinne kriterialer Validität als Anhaltspunkt dafür interpretieren, dass der LeHet-Musik-Test tatsächlich alltagsnahe bzw. praxisrelevante professionelle Kompetenzen von Musiklehrkräften misst – zumindest aber Fähigkeiten, über die im Beruf stehende Musiklehrkräfte in deutlich höherem Ausmaß verfügen als die studentischen Seminarteilnehmer*innen.

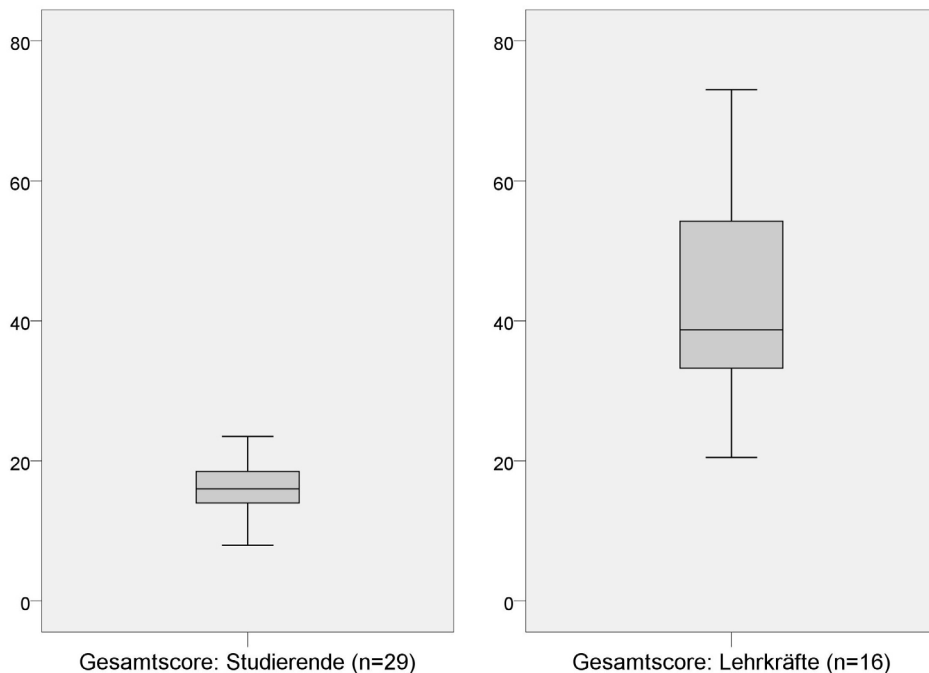


Abbildung 4: Testscores Studierende (2. Erhebung) und Lehrkräfte im Vergleich

Bezüglich der durchschnittlichen ‚Trefferquote‘ der Studierenden bei den 8 Items mit offenem Antwortformat (Hypothese 3) ergibt sich deskriptiv folgendes Bild:

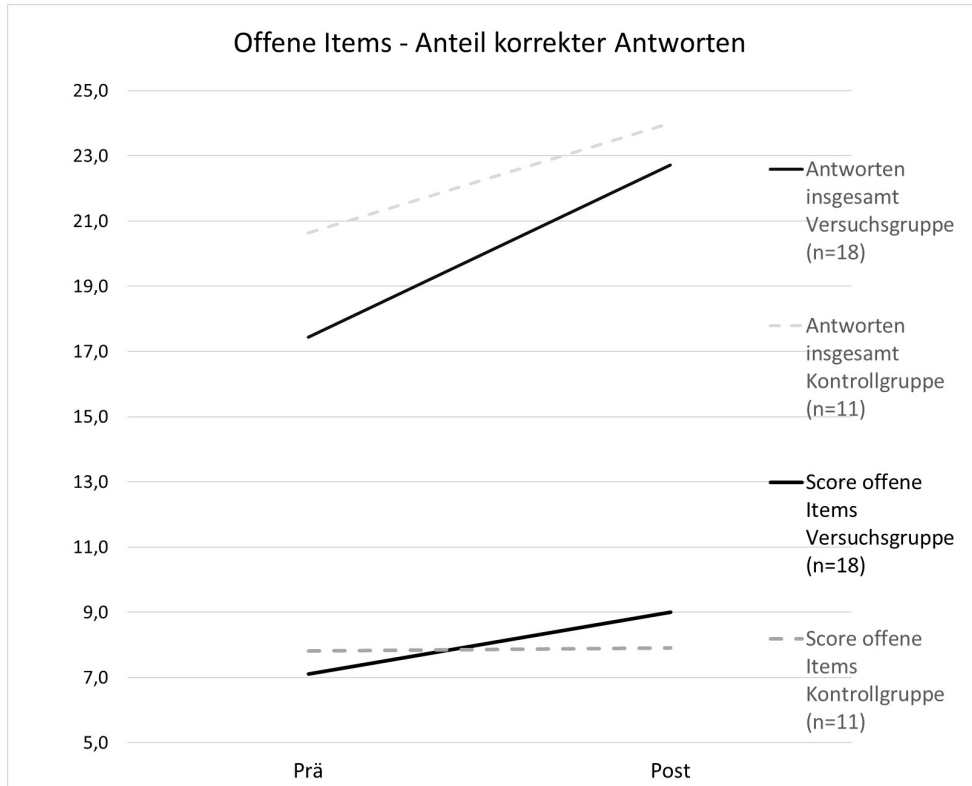


Abbildung 5: Anteil korrekter Antworten an den insgesamt gegebenen, substantiell verschiedenen Testantworten

Alle Studierenden brachten in der zweiten Erhebung deutlich mehr Antworten zu Papier als in der ersten (erste Erhebung: $M = 18.66$, $SD = 4.47$; zweite Erhebung: $M = 23.21$, $SD = 5.37$; $t(28) = 5.34$, $p < .001$ (zweiseitig), $d_{\text{repeated measures}} = 1.11$; 95% KI [0.556, 1.662]). Die Versuchsgruppe erzielte in den offenen Items auch höhere Testscores als in der ersten Erhebung (erste Erhebung: $M = 7.1$, $SD = 1.91$; zweite Erhebung: $M = 9.00$, $SD = 3.53$; $t(17) = 2.478$, $p = .02$, $d_{\text{repeated measures}} = .92$; 95% KI [0.23; 1.61]). Der Testscore der Versuchsgruppe änderte sich dagegen nicht signifikant. Die ‚Trefferquote‘ bei den Testantworten der Studierenden verbesserte sich hingegen nicht: Bei der Versuchsgruppe stagnierte sie bei ca. 40% (erste Erhebung: $M = 0.42$, $SD = 0.11$; zweite Erhebung: $M = 0.40$, $SD = 0.13$), bei der Kontrollgruppe verringerte sie sich sogar (erste Erhebung: $M = 0.39$, $SD = 0.14$; zweite Erhebung: $M = 0.31$, $SD = 0.14$). Hypothese 3 kann damit als

widerlegt gelten: Entgegen der Annahmen konnten die Studierenden der Versuchsgruppe zwar ihre Testleistungen verbessern; die höhere Zahl korrekter Testantworten in den offenen Items ging aber einher mit einer ebenfalls höheren Zahl falscher Vermutungen.

Dies lässt die Interpretation zu, dass der Erwerb von Professionswissen und professionellen Wahrnehmungsfähigkeiten nicht automatisch auch mit einer Verbesserung der Fähigkeit verbunden ist, beides gezielt und situationsgerecht einzusetzen. Angesichts der kleinen Stichprobe hat diese Annahme allerdings lediglich den Charakter einer empirisch gestützten Hypothese, die nach Validierung des Testinstruments in weiteren Studien zu prüfen wäre. Interessant ist auch in diesem Zusammenhang ein vergleichender Blick auf die Daten der 16 praxiserfahrenen Lehrkräfte, die am Test teilnahmen: Bei ihnen lag der Anteil korrekter Testantworten durchschnittlich bei 77% ($M = 0.77$, $SD = 0.10$) und damit erheblich höher als bei den Studierenden, was wieder im Einklang steht mit theoretischen Annahmen und empirischen Befunden zur Entwicklung professioneller Kompetenzen bei Lehrkräften (vgl. Dann & Haag, 2017, S. 108–109).

Fazit

Die Ergebnisse der vorliegenden Studie legen nahe, dass Studierende des nicht vertieften Lehramts Musik nach dem Besuch des neu entwickelten LeHet-Seminars über eine umfangreichere domänenspezifische Wissensbasis und über besser entwickelte situationsspezifische Fähigkeiten verfügen als vor Beginn. Deutliche Anhaltspunkte hierfür bieten die Ergebnisse eines entsprechenden vignettenbasierten Leistungstests, in dem problemhaltige Szenarien aus schulischem Singunterricht zu bearbeiten sind. Bei Studierenden einer Vergleichsgruppe ließ sich kein signifikanter Zuwachs in den interessierenden Kompetenzfacetten messen. Dies erlaubt den Schluss, dass die Leistungssteigerung der Studierenden des LeHet-Seminars tatsächlich durch die Teilnahme am Seminar beeinflusst wurde.

Durch die Teilnahme am Seminar verbesserte sich jedoch offenbar nicht die Fähigkeit der Studierenden, ihre Wissensbestände und situationsspezifischen *skills* gezielt und problemgerecht einzusetzen. Sie gaben auch nach der Teilnahme am Seminar durchschnittlich noch mehr falsche als korrekte Testantworten. Dieser Befund legt nahe, das Seminarkonzept zu überarbeiten: Die Studierenden sollten nicht nur neue Wissensbestände und diagnostische Fähigkeiten erwerben, sondern auch gezielt darin trainiert werden, beides zielgenau und reflektiert anzuwenden.

Berufserfahrene Musiklehrkräfte, die den LeHet-Musik-Test ebenfalls bearbeiteten, erzielten im Vergleich zu den getesteten Studierenden erheblich bessere Ergebnisse. Daraus lässt sich zunächst schließen, dass der Test tatsächlich professionelle Kompetenzen misst, die für das Unterrichten in der Domäne „Singen mit Schulklassen“ relevant sind. Darüber hinaus stellt sich aber die Frage, wie

die im Vergleich zu Lehrkräften schwachen Testleistungen der Studierenden einzuordnen sind. Gilt es in erster Linie, das Seminarkonzept noch einmal kritisch zu hinterfragen? Oder lässt sich aus solchen Ergebnissen der weitreichende Schluss ziehen, dass derart komplexe situationsspezifische Fähigkeiten im Rahmen eines universitären Seminars ohne unterrichtspraktische Anteile generell nur angebahnt, nicht aber in größerem Ausmaß trainiert werden können?

Um solchen Fragen nachgehen zu können, muss der Rahmen der vorliegenden Studie verlassen werden. Die schmale Datengrundlage lässt weder statistisch abgesicherte Aussagen über die Qualität des Testinstruments zu noch solche über die Gesamtpopulation ‚Studierende des Lehramts Musik‘ bzw. ‚Studierende des Lehramts Musik und berufserfahrene Musiklehrer*innen‘. Die Reichweite der hier vorgestellten Befunde ist limitiert auf die Stichprobe zum Erhebungszeitpunkt, was für den angestrebten Zweck einer Lehrveranstaltungsevaluation auch ausreicht.

Die Ergebnisse der Testauswertungen legen aber nahe, dass das Testkonzept „Vignettentest mit offenem Antwortformat“ grundsätzlich zur Messung professioneller Kompetenzen für die Domäne „Singen mit Schulklassen“ geeignet ist. Das neu entwickelte Instrument ermöglichte sowohl das Erfassen relativ kleinräumiger Veränderungen im Rahmen der Lehrveranstaltungsevaluation als auch den explorativen Vergleich zwischen Testleistungen der Studierenden und solchen berufserfahrener Musiklehrkräfte. Eine Validierung des LeHet-Musik-Tests an einer repräsentativen Stichprobe könnte somit lohnend sein, um ihn auch als Instrument zur wissenschaftlichen Grundlagenforschung einsetzen zu können – beispielsweise zur Bestimmung verschiedener Kompetenzniveaus oder zur längsschnittlichen Exploration der Frage, wann und wie sich professionelle Kompetenzen sowie die Fähigkeit, sie situationsgerecht einzusetzen, bei angehenden Musiklehrkräften entwickeln.

Bei einer Testvalidierung müsste sinnvollerweise nicht nur zwischen Studierenden und Lehrkräften differenziert werden, sondern auch zwischen Teilnehmer*innen aus vertieften, nicht vertieften und Nebenfachstudiengängen.¹⁰ Eine entsprechende A-priori-Analyse mit G*Power (Faul, Erdfelder, Lang & Buchner, 2007) ergibt für diesen Fall eine notwendige Stichprobenstärke von mindestens 216 (36 Teilnehmer*innen pro Analysezelle; $f = 0.25$, $\alpha \text{ err prob} = 0.05$, Power ($1 - \beta \text{ err prob}$) = 0.80).

Neben den vorgenommenen quantitativen Auswertungen erlaubt das offene Itemformat zudem qualitative Analysen der Testantworten, um beispielsweise Argumentationsstrukturen, Prä- und Fehlkonzepte von Studierenden im Kontext der Fallarbeit an Szenarien aus schulischem Musikunterricht zu explorieren.

Eine wichtige Limitierung betrifft die ökologische Validität des Tests. In Bezug auf universitäre Lehrveranstaltungen und deren angestrebtes Ergebnis darf dieser Anspruch wohl als eingelöst gelten. Mit Hilfe des Tests ließ sich dokumen-

10 Zur Begründung: vgl. z.B. Krauss et al., 2011, S. 152.

tieren, dass bei der Versuchsgruppe offensichtlich im Bereich professioneller Dispositionen und Wahrnehmungsfähigkeiten aus der ‚Außensicht‘ positive Veränderungen stattgefunden hatten. Das Notieren von Diagnosen und Handlungsvorschlägen als Reaktion auf die Konfrontation mit problemhaltigen Szenarien aus schulischem Musikunterricht ist aber ein Akt der Performanz, der sich deutlich vom „Handeln unter Druck“ (Dann & Haag, 2017, S. 97) im Unterricht unterscheidet. Wie für etliche andere Tests auch bleibt im vorliegenden Fall offen, in welchem Bezug die Testleistungen zu tatsächlichem unterrichtlichem Handeln, zu Unterrichtsqualität und zu Lernerfolgen von Schüler*innen stehen. Die Entwicklung von Untersuchungsansätzen und -instrumenten für den Bereich des Singens im schulischen Musikunterricht, mit deren Hilfe sich solche Zusammenhänge im Rahmen von mehrperspektivischen Designs untersuchen lassen, stellt ein Desiderat für die kommenden Jahre dar.

Literatur

- Arnold-Joppich, H. (2015). Umgang mit der Kinderstimme – „Stimm-Bildung“. In M. Fuchs (Hrsg.), *Musikdidaktik Grundschule. Theoretische Grundlagen und Praxisvorschläge* (S. 106–119). Innsbruck: Helbling.
- Baron-Boldt, J., Schuler, H. & Funke, U. (1988). Prädiktive Validität von Schulabschlussnoten: Eine Metaanalyse. *Zeitschrift für pädagogische Psychologie*, 2(2), 79–90.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst. (2014). *LehrplanPLUS Grundschule. Lehrplan für die bayerische Grundschule*. Verfügbar unter: <https://www.lehrplanplus.bayern.de/schulart/grundschule> [27.12.2019].
- Berliner, D. C. (2005). The Near Impossibility of Testing for Teacher Quality. *Journal of Teacher Education*, 56(3), 205–213. <https://doi.org/10.1177/0022487105275904>
- Blömeke, S., Bremerich-Vos, A., Haudeck, H., Kaiser, G., Nold, G., Schwippert, K. et al. (Hrsg.). (2011). *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT*. Münster: Waxmann.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Bromme, R. (1992). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens* (Huber-Psychologie-Forschung). Bern: Huber.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert, N. Birbaumer & C. F. Graumann (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie / Themenbereich D; Ser. 1; Bd. 3, S. 177–212). Göttingen: Hogrefe.
- Dann, H.-D. & Haag, L. (2017). Lehrerkognitionen und Handlungsentscheidungen. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungspers-*

- pektiven und methodische Zugänge (3. Auflage, Schule und Gesellschaft, Bd. 24, S. 89–120). Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-15083-9>
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175–191.
- Fuchs, M. (2015). Singen im Musikunterricht. In M. Fuchs (Hrsg.), *Musikdidaktik Grundschule. Theoretische Grundlagen und Praxisvorschläge* (S. 120–141). Innsbruck: Helbling.
- Göllner, M. & Niessen, A. (2016). Planungsanpassung als adaptive Maßnahme in musikpädagogischen Lernsituationen im Spiegel qualitativer Interviews. In J. Knigge & A. Niessen (Hrsg.), *Musikpädagogik und Erziehungswissenschaft. Music Education and Educational Science* (Musikpädagogische Forschung, Bd. 37, S. 121–135). Münster: Waxmann.
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E. & Williamson, P. (2009). Teaching Practice: A Cross-Professional Perspective. *Teachers College Record*, 111(9), 2055–2100. Verfügbar unter: <http://www.tcrecord.org/Content.asp?ContentId=15018> [23.11.2020].
- Hammann, M. & Jördens, J. (2014). Offene Aufgaben codieren. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 169–178). Berlin: Springer Spektrum.
- Heinzel, F. & Krasemann, B. (2014). Lehrerbildung mit dem Online-Fallarchiv Schulpädagogik. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba & J. Wildt (Hrsg.), *Neues Handbuch Hochschullehre* (S. 43–68). Berlin: DUZ Verlags- und Medienhaus.
- Hellermann, C., Gold, B. & Holodynski, M. (2015). Förderung von Klassenführungsfähigkeiten im Lehramtsstudium. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*. <https://doi.org/10.1026/0049-8637/a000129>
- Henning, H. (2014). *Qualität in der vokalpädagogischen Praxis mit Kindern im Grundschulalter. Impulse zur Qualitätsentwicklung*. Dissertation, Hochschule für Musik. Würzburg. Verfügbar unter: https://opus.bibliothek.uni-wuerzburg.de/opus4-wuerzburg/frontdoor/deliver/index/docId/13022/file/Henning_Dissertation_Vokalpaedagogische_Praxis.pdf [10.10.2019].
- Hofmann, B. (2015). Musik machen mit der Stimme: Vokaldidaktik. In M. Loritz & C. Schott (Hrsg.), *Musik – Didaktik für die Grundschule* (Didaktik für die Grundschule, S. 105–118). Berlin: Cornelsen.
- Klusmann, U. (2011). Individuelle Voraussetzungen von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 297–304). Münster: Waxmann.
- Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M. et al. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 135–161). Münster: Waxmann.
- Krauss, S., Lindl, A. & Schilcher, A. (Hrsg.). (2017). *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik*. Münster: Waxmann.

- Krauss, S., Lindl, A., Schilcher, A. & Tepner, O. (2017). Das Forschungsprojekt FALKO – ein einleitender Überblick. In S. Krauss, A. Lindl & A. Schilcher (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 9–65). Münster: Waxmann.
- Kunina-Habenicht, O., Schulze-Stocker, F., Kunter, M., Baumert, J., Leutner, D., Förster, D. et al. (2013). Die Bedeutung der Lerngelegenheiten im Lehramtsstudium und deren individuelle Nutzung für den Aufbau des bildungswissenschaftlichen Wissens. *Zeitschrift für Pädagogik*, 59(1), 1–23. <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-119245>
- Maschke, H. (2008). *Spaß beim Singen. Lieder für viele Gelegenheiten* (Lehrer-Bücherei Grundschule). Berlin: Cornelsen Scriptor.
- Mayer, J. & Wellnitz, N. (2014). Die Entwicklung von Kompetenzstrukturmodellen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschafts-didaktischen Forschung* (S. 19–29). Berlin: Springer Spektrum.
- Mohr, A. (2013). *Handbuch der Kinderstimmbildung* (6. Auflage, Studienbuch Musik). Mainz: Schott.
- Morris, S. B. & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105–125. <https://doi.org/10.1037/1082-989X.7.1.105>
- Mulder, R. H. & Gruber, H. (2011). Die Lehrperson im Lichte von Professions-, Kompetenz- und Expertiseforschung – die drei Seiten einer Medaille. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen empirischer Bildungsforschung. Traditionslinien und Perspektiven [Klaus Beck zum 70. Geburtstag gewidmet]* (S. 427–438). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Puffer, G. & Hofmann, B. (2017). FALKO-M: Entwicklung und Validierung eines Testinstruments zum domänenspezifischen Professionswissen von Musiklehrkräften. In S. Krauss, A. Lindl & A. Schilcher (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 245–289). Münster: Waxmann.
- Seidel, T., Blomberg, G. & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. *Zeitschrift für Pädagogik*, Beiheft, 56, 296–306. <http://nbn-resolving.org/urn:nbn:de:0111-opus-34384>
- Seidel, T. & Thiel, F. (2017). Standards und Trends der videobasierten Lehr-Lernforschung. *Zeitschrift für Erziehungswissenschaft*, 20(S1), 1–21. <https://doi.org/10.1007/s11618-017-0726-6>
- Sherin, M. G. (2007). The development of teachers' professional vision in video clubs. In R. Goldman, R. Pea, B. Barron & S. J. Derry (Hrsg.), *Video Research in the Learning Sciences* (S. 383–395). Hoboken: Taylor and Francis.
- Sherin, M. G. & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60(1), 20–37. <https://doi.org/10.1177/0022487108328155>
- Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK). (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. Beschluss der KMK vom 16.10.2008 i. d. F.*

- vom 2.10.2017. Verfügbar unter: http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf [10.10.2019].
- Universität Augsburg. (2013). Prüfungsordnung der Universität Augsburg für die Modulprüfungen im Rahmen der Ersten Prüfung für ein Lehramt an öffentlichen Schulen und für die Lehramtsbezogenen Bachelorstudiengänge der Universität Augsburg. LPO-UA. Verfügbar unter: https://assets.uni-augsburg.de/media/filer_public/e8/8e/e88e3d63-62e1-4e41-947b-edcb6b311e1e/1-174-2-004.pdf [31.12.2020].
- Wieblitz, C. (2013). *Lebendiger Kinderchor. Kreativ, spielerisch, tänzerisch; Anregungen und Modelle* (5. Auflage). Boppard am Rhein: Fidula-Verlag.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.

Gabriele Puffer
Universität Augsburg
Lehrstuhl für Musikpädagogik
Universitätsstr. 26
86159 Augsburg
gabriele.puffer@phil.uni-augsburg.de