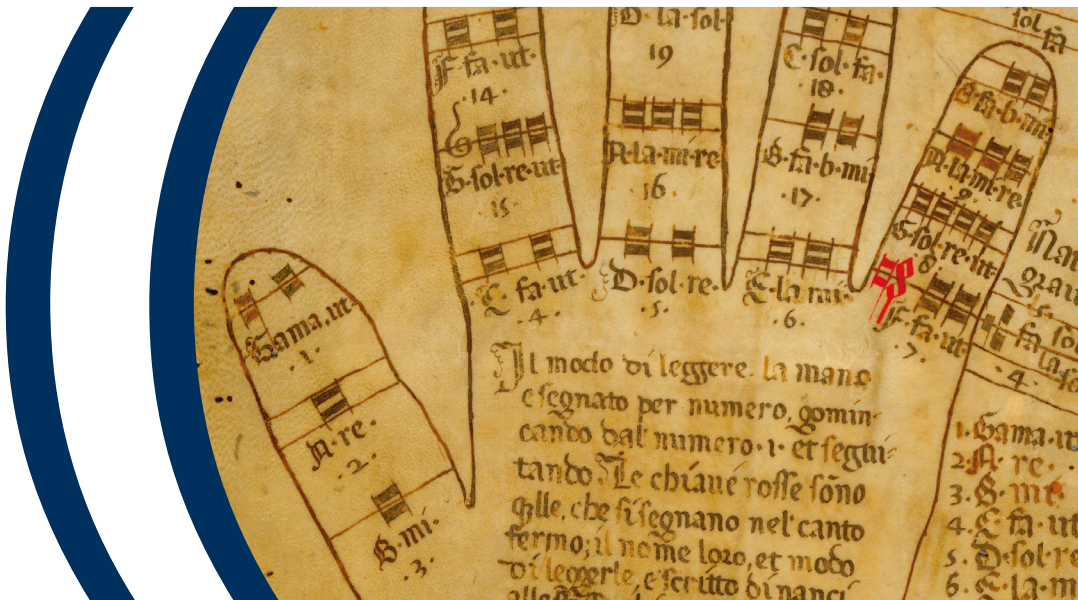


Mario Frei

Erklären im Musikunterricht

Eine Studie zu Qualitätsmerkmalen



Perspektiven musikpädagogischer Forschung

herausgegeben von
Prof. Dr. Jens Knigge
Prof. Dr. Ulrike Kranefeld
Prof. Dr. Anne Niessen
Prof. Dr. Christine Stöger

Band 19

Mario Frei

Erklären im Musikunterricht

Eine Studie zu Qualitätsmerkmalen



Waxmann 2023
Münster • New York

Gefördert durch die Gisela und Peter W. Schatt Stiftung

GISELA^{UND}
PETER W.
SCHATT
STIFTUNG

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.

Perspektiven musikpädagogischer Forschung, Band 19

ISSN 2198-1973

Print-ISBN 978-3-8309-4772-1

E-Book-ISBN 978-3-8309-9772-6

<https://doi.org/10.31244/9783830997726>

Waxmann Verlag GmbH, 2023

www.waxmann.com

info@waxmann.com

Umschlaggestaltung: Anne Breitenbach, Münster

Titelbild: Guidonian hand

Dieses Werk ist unter der Lizenz CC BY-NC-SA 4.0 veröffentlicht:

Namensnennung – Nicht-kommerziell –

Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-NC-SA 4.0)



Diese Lizenz gilt nur für das Originalmaterial. Alle gekennzeichneten Fremdinhalte (z.B. Abbildungen, Fotos, Zitate etc.) sind von der CC-Lizenz ausgenommen und für deren Wiederverwendung ist es ggf. erforderlich, weitere Nutzungsgenehmigungen beim jeweiligen Rechteinhaber einzuholen.

Für Marianna, Luca & Béla

Dank

Am Gelingen dieser Arbeit haben viele wundervolle Menschen einen großen Anteil. Es ist ein unglaubliches Gefühl, euch alle an meiner Seite zu wissen.

Zunächst gilt mein Dank meinem Doktorvater Bernhard Hofmann. Lieber Bernhard, ich möchte dir von Herzen für deine in jeder Phase wertschätzende wie motivierende Betreuung und deine unzählbaren wertvollen Anregungen danken. Danke für deine Bereitschaft, auf meinen fahrenden Zug aufzuspringen und ihn wieder aufs Gleis zu setzen.

Genauso möchte ich Sven Hilbert danken, der die Arbeit von Beginn an methodisch betreut hat. Lieber Sven, vielen Dank für dein stetes Vertrauen, deine Geduld und deine Zuversicht, mit der du mir ein (damals noch) unbekanntes Arbeitsfeld nähergebracht hast, das ich heute nicht mehr missen möchte. Danke, dass ich durch dich erfahren durfte, wie erfüllend das wissenschaftliche Arbeiten sein kann.

Außerdem gilt mein Dank Gabriele Puffer. Liebe Gaby, du hast mich nicht nur mit der Lust an musikpädagogischer Forschung infiziert, sondern mir dabei auch auf unterschiedlicher Ebene so viel beigebracht. Danke.

Dann danke ich meinem Kollegen Alfred Lindl. Lieber Alfred, ich danke dir nicht nur für die langjährige, wertschätzende und verlässliche Zusammenarbeit, sondern auch für dein stets offenes Ohr, die inspirierenden fachlichen Diskussionen und deine wertvollen Empfehlungen bei statistischen Analysen.

Ein großes Dankeschön geht an die genialste Arbeitsgruppe. Ich danke allen für die gegenseitige Unterstützung und den großartigen Zusammenhalt im Team.

Dann möchte ich mich sehr bei Anita Schilcher und Stefan Krauss bedanken, die das FALKE-q-Projekt geleitet haben, in dessen Rahmen die Arbeit entstand. Liebe Anita, lieber Stefan, vielen Dank für alles, was ich in den Jahren von euch lernen durfte, vielen Dank für eure Unterstützung und euer außerordentliches Engagement. Ihr habt das FALKE-q-Projekt zu etwas ganz Besonderem gemacht.

Auch gilt ein besonderer Dank Karsten Rincke. Lieber Karsten, vielen Dank für deine Hilfe beim Aufzeigen neuer Wege und deine bemerkenswerte Unterstützung bei der Umsetzung.

Ein großes DANKE geht an alle Falken. Liebe Lisa, liebe Katharina, liebe Christina, liebe Maria, liebe Anna, liebe Eileen, liebe Renate, liebe Simone, liebe Jana, lieber Matthias, lieber Michael, danke, dass ich Teil von einem so großartigen Team sein durfte. Danke für die immer wertschätzende und rücksichtsvolle Zusammenarbeit. Danke für die vielen neuen Perspektiven, die ich einnehmen konnte.

Vielen Dank an meinen Kollegen Patrick Ehrich für die geduldige und professionelle Produktion der Videovignetten.

Ein großer Dank geht an alle Hilfskräfte, die mich während der Projektlaufzeit auf unterschiedliche Weise unterstützt haben, insbesondere Laura Scharl für das umfangreiche Codieren der offenen Antworten.

Danke an alle Schüler*innen, Studierenden, Lehrkräfte und Didaktiker*innen, die mit großem Zeitaufwand an der Studie teilgenommen haben.

Vielen Dank an Bernhard Schmidtner für das schnelle, aufmerksame und gründliche Korrekturlesen.

Von Herzen möchte ich meinen Eltern und meinem Bruder für die Unterstützung über mein ganzes Leben hinweg bedanken. Ihr steht immer hinter mir. Danke.

Zu guter Letzt gilt mein größter Dank meiner wunderbaren Frau und meinen beiden wundervollen Kindern. Liebe Marianna, ohne deine bedingungslose Unterstützung und deinen Glauben an mich wäre diese Arbeit nicht möglich gewesen. Du hast einen sehr großen Anteil. Lieber Luca, lieber Béla, danke, dass ihr so geduldig mit mir wart und viele Wochenenden auf mich verzichtet habt. Ihr zeigt mir jeden Tag aufs Neue, was wirklich wichtig ist im Leben.

Bamberg, im Oktober 2023

Mario Frei

Inhalt

1	Einleitung	13
2	Erklären	18
2.1	Zum Begriff Erklären	18
2.2	Bildungswissenschaftlicher Diskurs	23
2.2.1	Unterrichtsqualität	24
2.2.2	Professionelle Kompetenzen von Lehrkräften	46
2.3	Erklären im Unterricht	60
2.3.1	Charakteristika	61
2.3.2	Qualitätsmerkmale	68
2.4	Erklären im Musikunterricht	86
2.4.1	Eigenschaften von Musikunterricht	87
2.4.2	Erklären und Instruieren	89
2.4.3	Erklären musiktheoretischer Inhalte	104
3	Zielsetzung und Fragestellungen	114
3.1	Zielsetzung	114
3.2	Forschungsfragen und Hypothesen	116
3.2.1	Zusammensetzung der Statusgruppen	117
3.2.2	Unterschiede in der Erklärqualität bezüglich Erklärmodus und Statusgruppe	117
3.2.3	Zusammenhang zwischen Erklärqualität und Qualitätsmerkmalen	120
3.2.4	Zusammenhang zwischen Vorwissen der Schüler*innen und Erklärqualität	121
4	Methode	122
4.1	Forschungsmethodische Überlegungen	122
4.1.1	Perspektiventriangulation	122
4.1.2	Methodentriangulation	123
4.2	Forschungsdesign und Stichprobe	124
4.2.1	Forschungsdesign	124
4.2.2	Stichprobe	125
4.3	Videovignetten	127
4.3.1	Erklärgegenstände	128
4.3.2	Gestaltungsvorgaben	130
4.3.3	Entstehung und Qualitätssicherung	131
4.4	Messinstrumente	131
4.4.1	Onlinefragebogen	131
4.4.2	Wissenstest	134
4.5	Operationalisierung des Messgegenstands	135

4.5.1	Sprech- und Körperausdruck	136
4.5.2	Einsatz musikbezogener Repräsentationen	136
4.5.3	Sprachliche Verständlichkeit	139
4.5.4	Strukturiiertheit	140
4.5.5	Adressatenorientierung	141
4.6	Weitere erhobene Variablen	142
4.6.1	Persönlichkeitswirkung	142
4.6.2	Zusätzliche Kontrollvariablen	142
4.6.3	Offene Itemformate	144
4.7	Pilotierung des Erhebungsinstruments	144
4.8	Durchführung der Erhebung	146
4.9	Analyseverfahren	147
4.9.1	Statistische Analyse	147
4.9.2	Qualitative Inhaltsanalyse	150
5	Ergebnisse	159
5.1	Vorbereitende Analyse zur Zusammensetzung der Statusgruppen	159
5.2	Gütekriterien	162
5.2.1	Psychometrische Gütekriterien der geschlossenen Fragebogenitems	162
5.2.2	Psychometrische Gütekriterien des Wissenstests	169
5.2.3	Spezifische Gütekriterien der offenen Fragebogenitems	170
5.3	Unterschiede in der Bewertung der Erklärqualität	176
5.3.1	Varianzanalyse mit gemischtem Design	177
5.3.2	Faktor Statusgruppe	185
5.3.3	Faktor Erklärmodus	195
5.3.4	Zusammenfassung	204
5.4	Qualitätsmerkmale unterrichtlichen Erklärens im Fach Musik	206
5.4.1	Zusammenhang zwischen Erklärqualität und ausgewählten Merkmalen	206
5.4.2	Subjektive Theorien zu den ausgewählten Merkmalen	221
5.4.3	Subjektive Theorien zu weiteren Merkmalen	236
5.4.4	Zusammenfassung	239
5.5	Zusammenhang zwischen Vorwissen der Schüler*innen und Erklärqualität	243
6	Diskussion	248
6.1	Zusammenfassung und Einordnung in den Forschungsdiskurs	248
6.1.1	Erklärqualität	248
6.1.2	Qualitätsmerkmale	250
6.2	Einschränkungen und Ausblick	255

6.2.1	Forschungsmethodisches Vorgehen	255
6.2.2	Erklären im Musikunterricht	257
6.3	Fazit	258
Abbildungsverzeichnis		261
Tabellenverzeichnis		262
Literatur		266

Der Anhang steht zum freien Download unter www.waxmann.com/buch4772 zur Verfügung.

1 Einleitung

„Sie gehören zum alltäglichen Geschäft der Lehrperson:
Gute Erklärungen sind im Unterricht unverzichtbar.“
(Pauli, 2015, S. 44)

Paulis Aussage stellt zwei Thesen auf: Erklären gilt zum einen fachübergreifend als wesentliche Unterrichtshandlung einer Lehrkraft und Unterricht vermag zum anderen ohne gute Erklärungen nicht auszukommen. Hervor treten zudem die zwei bildungswissenschaftlichen Diskurse um professionelle Kompetenz von Lehrkräften und Unterrichtsqualität, welche die Frage nach gutem Erklären grundieren und für die vorliegende Arbeit leitend sind. Allgemein wird das Erklären in der Unterrichtsforschung zu den Kerntätigkeiten einer Lehrkraft gezählt (z. B. Charalambous et al., 2011; Leinhardt, 2001; Pauli, 2015). Empirische Befunde zeigen dabei, dass Schüler*innen auf die Frage nach den entscheidenden Fähigkeiten einer Lehrkraft häufig antworten, sie müsse gut erklären können (z. B. Wörn, 2014; Wragg & Wood, 1984). Im Kontext der Lehrkräftebildung zählt Erklären zu den *core practices* und gilt dementsprechend als eine im Unterricht häufig vorkommende, erlernbare und fachübergreifend bedeutsame Praktik einer Lehrkraft (Fraefel & Scheidig, 2018; Grossman et al., 2009).

Demnach wäre einerseits zu erwarten, dass die These der *Unverzichtbarkeit guten Erklärens* durch theoretische Modelle und empirische Erkenntnisse hinreichend fundiert und belegt ist. Es zeigt sich jedoch, dass breit rezipierte Handbücher der Lehr- und Lernforschung (z. B. Kunter & Trautwein, 2013; Terhart et al., 2014) das Erklären nahezu ausblenden – es existieren lediglich vereinzelte theoretische Modellierungen (z. B. Kiel, 1999). Auch liegen zwar erste empirische Erkenntnisse in einzelnen Fachdidaktiken vor (z. B. Findeisen, 2017; Kulgemeyer & Tomczyszyn, 2015; Neumeister, 2011; Rathausky, 2010; Wörn, 2014), allerdings werden diese weitgehend losgelöst von einem fachübergreifenden Ansatz diskutiert, der in der Unterrichtsforschung zunehmend gefordert wird (Praetorius, Herrmann et al., 2020).

Andererseits provoziert das obenstehende Zitat aus musikpädagogischer Perspektive grundsätzliche Fragen hinsichtlich Stellung, Funktion und Qualität des Erklärens von Lehrkräften¹ im Kontext von Musikunterricht. Während dazu in den Bereichen Wirtschaft, Naturwissenschaft, (Fremd-)Sprachen und Mathematik bereits erste Erkenntnisse vorliegen (z. B. Findeisen, 2017; Kulgemeyer & Tomczyszyn, 2015; Neumeister, 2011; Rathausky, 2010; Wörn, 2014), zeigt sich die Forschungslage in der Musikpädagogik defizitär. Zwar liegen erste Ansätze einer theoretischen Modellierung guten Erklärens vor (Frei et al., 2022), allerdings fehlen insbesondere empirische Befunde zu Qualitätsmerkmalen des Erklärens im Musikunterricht (Frei, 2019, S. 124). Auch findet sich das Wort *Erklären*, verstanden als Unterrichtshandeln einer Lehrkraft, nur sehr vereinzelt und wird

1 Wenn im Folgenden der vorliegenden Arbeit von Erklären die Rede ist, meint dies stets das Erklären von Lehrkräften; andernfalls wird explizit darauf hingewiesen. Zum Erklären von Schüler*innen, das nicht im Interesse der vorliegenden Arbeit steht, siehe z. B. Kulgemeyer & Schecker, 2013.

dabei weder ausführlich thematisiert noch als expliziter Forschungsgegenstand verstanden (z. B. Eberhard, 2010, S. 135; Fuchs, 2015, S. 102; Puffer, 2019, S. 229).²

Generell sind musikdidaktische Forschungsbemühungen zu den zwei im einleitenden Zitat angesprochenen und eng miteinander verknüpften Bereichen von Unterrichtsforschung erst in Ansätzen zu beobachten. Forschung über die professionelle Kompetenz von Musiklehrkräften steht erst seit einigen Jahren im Fokus der Aufmerksamkeit (Lehmann-Wermser & Krause-Benz, 2013, S. 8; Lessing & Stöger, 2018, S. 132–138), generiert seitdem jedoch vermehrt empirische Befunde zum Professionswissen oder zu affektiv-motivationalen Komponenten professioneller Kompetenz (z. B. Harnischmacher et al., 2017; Puffer & Hofmann, 2017; für eine zusammenfassende Modellierung professioneller Kompetenz von Musiklehrkräften siehe Puffer, 2021). Trotz einzelner Beiträge in den letzten zwei Jahrzehnten resümiert Kranefeld (2021), dass der Musikpädagogik „theoretisch ausgearbeitete und empirisch fundierte Modellierungen zur fachbezogenen Unterrichtsqualität“ (S. 2) fehlen.

Die Frage nach Stellung, Funktion und Qualität des Erklärens im Musikunterricht verortet sich an der Schnittstelle zwischen allgemeiner und musikpädagogischer Unterrichtsforschung. Antworten auf diese Fragen sind demnach im Spannungsfeld von Generik und Fachspezifik zu suchen (Praetorius, Herrmann et al., 2020; für den musikpädagogischen Diskurs siehe Kranefeld, 2021). Die vorliegende Untersuchung rückt das Erklären als Unterrichtshandeln einer Musiklehrkraft ins Zentrum des Forschungsinteresses. Sie möchte Erkenntnisse zu Merkmalen guten Erklärens im Musikunterricht gewinnen, indem sie Beurteilungen auswertet, die aus verschiedenen Beobachtungsperspektiven gegeben werden. Außerdem soll die Arbeit auch einen Beitrag zum Stand der Diskussion um Unterrichtsqualität aus Perspektive der Musikdidaktik³ leisten.

Mit der Inanspruchnahme des Begriffs *gut* folgt die Arbeit einer Klassifikation von Unterrichten, der sich nicht nur die allgemeine (z. B. Kunter & Trautwein, 2013), sondern auch die musikpädagogische Unterrichtsforschung (z. B. Kranefeld, 2021; Puffer, 2021) bedient: Berliner (2005) spricht von *gutem* Unterrichten, wenn dabei normative und pädagogische Prinzipien eingehalten werden; führt das Unterrichtshandeln zum Erreichen der Lernziele und kann damit dessen Wirksamkeit nachgewiesen werden, wäre das Unterrichten *effektiv*; von *qualitätvollem* Unterrichten kann hingegen erst die Rede sein, wenn beide Aspekte zutreffen.⁴

2 Ob und inwieweit dieser Befund mit musikästhetischen Auffassungen zusammenhängt, nach welchen sich „Musik selbst erklärt“ (Kühn, 2006, S. 25), wird im Rahmen dieser Arbeit nicht untersucht.

3 Der Begriff *Musikdidaktik* wird uneinheitlich verwendet (Jank et al., 2020; anders: Hörmann & Meidel, 2016). Im Folgenden wird darunter ein Teilgebiet der Musikpädagogik verstanden, welches sich als wissenschaftliche Disziplin mit fachdidaktischer Forschung zum allgemeinbildenden schulischen Musikunterricht befasst und sich weniger auf musikpädagogische Aktivitäten außerhalb des schulischen Rahmens bezieht, wie etwa Instrumental- und Gesangsunterricht oder Musikgeragogik (die Definition orientiert sich an Kranefeld, 2021).

4 Um einen Anschluss an den Diskurs um Unterrichtsqualität herzustellen, bei dem diese sprachlichen Nuancierungen teilweise nicht vorgenommen werden (z. B. Helmke, 2017), wird im Folgenden auch der Begriff Erklärqualität verwendet und soll stets – außer es ist im Text explizit anders angemerkt – als gutes Erklären im Sinne Berliners (2005) verstanden werden. Demnach sind die Termini

Ziele der Arbeit

Unter Bezugnahme auf die formulierten Desiderata verfolgt diese Arbeit im Wesentlichen fünf Ziele:

- Orientiert am Diskurs um Unterrichtsqualität sollen Merkmale guten Erklärens unter Rekurs auf fachübergreifende Erkenntnisse identifiziert und für den Musikunterricht konkretisiert werden.
- Es soll ein Instrument entwickelt werden, das zum einen verschiedene Perspektiven auf Erklären abbildet und zum anderen nicht nur die Erklärqualität, sondern auch Beurteilungen zu ausgewählten Qualitätsmerkmalen erfassen kann.
- Auf Basis dessen sollen Unterschiede in der Erklärqualität unter Berücksichtigung der Zugehörigkeit zur Statusgruppe (Schüler*innen, Studierende und Lehrende) und dem Einsatz optischer oder akustischer Repräsentationsformen analysiert werden.
- Weiterhin soll für jede Statusgruppe getrennt untersucht werden, welche Merkmale für die Bewertung von Erklärqualität Bedeutung haben.
- Ergänzend soll für die Gruppe der Schüler*innen untersucht werden, inwieweit ihr Vorwissen Bedeutung für die Bewertung der Erklärqualität einnimmt.

Aufbau der Arbeit

Die vorliegende Arbeit setzt sich aus einem theoretischen und einem empirischen Teil zusammen und folgt dem nachstehenden inhaltlichen Aufbau:

Kapitel 2 entfaltet das Konstrukt Erklären und grenzt den Gebrauch des Begriffs zunächst im Kontext von Unterricht ein (2.1), bevor die zugrunde liegenden bildungswissenschaftlichen Diskurse über Unterrichtsqualität (2.2.1) und Erklären als professionelle Kompetenz von Lehrkräften (2.2.2) dargestellt werden. Mit Blick auf unterrichtliches Erklären liegt der Fokus neben der Zusammenstellung von Charakteristika (2.3.1) insbesondere auf der Identifikation von Qualitätsmerkmalen (2.3.2). Im letzten Abschnitt des Theorieteils erfolgt eine Einordnung von Erklären im Kontext von Musikunterricht. Vor dem Hintergrund der zunächst vornehmlich fachübergreifenden Perspektive werden zunächst zentrale Eigenschaften von Musikunterricht (2.4.1) entfaltet; dies dient sowohl der Einordnung fachübergreifender als auch der Rahmung musikspezifischer Aspekte unterrichtlichen Erklärens. Auf Basis dessen erfolgt die Modellierung von Erklären im Musikunterricht anhand einer Auseinandersetzung zwischen Erklären und Instruieren (2.4.2) als zwei naheverwandte Unterrichtshandlungen einer Musiklehrkraft. Als besonders *erklärbedürftige* Gegenstände des Musikunterrichts werden musiktheoretische Inhalte (2.4.3) aus dem Bereich der Allgemeinen Musiklehre dargestellt.

Kapitel 3 formuliert daraufhin die Zielsetzung der vorliegenden Arbeit (3.1). Hinsichtlich der Darstellung von Fragestellungen, die teils explorativen Charakter aufweisen,

Merkmale von Unterrichtsqualität und *Merkmale guten Unterrichts* ebenso synonym zu verstehen wie *Qualitätsmerkmale von Erklären* und *Merkmale guten Erklärens*.

teils jedoch auch mit Hypothesen einhergehen, werden zunächst vorbereitende Analysen zur Zusammensetzung der Statusgruppen angestellt (3.2.1). Anschließend werden Fragestellungen und Hypothesen im Kontext möglicher Unterschiede in der Bewertung der Erklärqualität (3.2.2) sowie Zusammenhänge zwischen Erklärqualität und ausgewählten Merkmalen (3.2.3) entfaltet. Abschließend wird eine Hypothese über den Zusammenhang zwischen dem Vorwissen der Schüler*innen und deren Bewertung der Erklärqualität aufgestellt (3.2.4).

Den empirischen Teil eröffnet *Kapitel 4*, das die Methode dieser Untersuchung erläutert. Dabei werden zunächst forschungsmethodische Überlegungen zu Perspektiven (4.1.1) und Methodentriangulation (4.1.2) angestellt, bevor das Forschungsdesign (4.2.1) und die Stichprobe (4.2.2) der Studie vorgestellt werden. Die Erläuterung der als Stimuli eingesetzten Videovignetten umfasst die Darstellung der erklärten Gegenstände (4.3.1), die bei der Erstellung berücksichtigten Gestaltungsvorgaben (4.3.2) sowie die Mechanismen zur Qualitätssicherung (4.3.3). Als zentrales Messinstrument der Studie dient ein Onlinefragebogen (4.4.1), der anhand geschlossener und offener Antwortformate (für die Gruppen der Studierenden und Lehrenden) insbesondere Bewertungen zu Qualität und Beurteilungen zu ausgewählten Qualitätsmerkmalen unterrichtlichen Erklärens messen kann. Darüber hinaus werden die kurzen Wissenstests (4.4.2) eingeführt, mit denen das Vorwissen der Schüler*innen zu den erklärten Themen erhoben und potenzielle Lernzuwächse nach den Erklärungen gemessen werden sollen. Vor diesem Hintergrund wird die Operationalisierung des Messgegenstands (4.5) für jedes Merkmal separat dargelegt und um weitere Kontrollvariablen sowie die Items mit offenen Antwortformaten ergänzt (4.6). Aufgrund der Neukonstruktion wird anschließend die ausführliche Pilotierung des Erhebungsinstruments (4.7) beziehungsweise der darin implementierten Items erläutert, bevor die Durchführung der Erhebung (4.8) dargestellt wird. Abschließend erfolgt der Bericht der statistischen (4.9.1) sowie inhaltsanalytischen Verfahren (4.9.2), die bei der Datenauswertung zum Einsatz gekommen sind.

Kapitel 5 legt die empirischen Befunde dar und gibt Antworten auf die in Kapitel 3 formulierten Fragestellungen und Hypothesen. Dabei werden zunächst die Ergebnisse der vorbereitenden Analysen über die Zusammensetzung der Statusgruppen berichtet (5.1). Es folgen die Darstellungen von psychometrischen Gütekriterien der geschlossenen und quantitativ ausgewerteten Fragebogenitems (5.2.1) sowie von spezifischen Gütekriterien der offenen und qualitativ ausgewerteten Antworten (5.2.2). Unterschiede in der Bewertung der Erklärqualität werden anhand von Varianzanalysen mit gemischtem Design (5.3.1) unter Berücksichtigung der Faktoren Statusgruppe (5.3.2) und Erklärmodus (5.3.3) aufgezeigt. Anschließend erfolgt die Darstellung der Ergebnisse aus Regressionsanalysen, in denen Zusammenhänge zwischen Erklärqualität und ausgewählten Qualitätsmerkmalen getrennt nach Statusgruppen für Schüler*innen, Studierende und die Gruppe der Lehrenden berichtet werden (5.4.1). Es folgt die Analyse der offenen Antworten für Studierende und Lehrende, wobei zunächst deren subjektive Theorien zu den fünf ausgewählten Qualitätsmerkmalen (5.4.2) berichtet werden, bevor diese um weitere potenzielle Qualitätsmerkmale ergänzt werden (5.4.3). Abschließend wird beantwortet, inwieweit

das Niveau des Vorwissens der Schüler*innen mit deren Bewertung der Erklärqualität zusammenhängt (5.5).

Kapitel 6 diskutiert die Ergebnisse. Dazu werden die zentralen Befunde zusammengefasst und in den musikpädagogischen sowie bildungswissenschaftlichen Forschungskontext eingeordnet (6.1). Nachdem das forschungsmethodische Vorgehen kritisch reflektiert und Grenzen der Aussagekraft benannt wurden, erfolgt ein Ausblick, der offene Desiderata aufzeigt und zukünftige Forschungsperspektiven skizziert (6.2). Das abschließende Fazit geht auf die praktische Relevanz der Ergebnisse ein (6.3).

2 Erklären

Der Begriff *Erklären* begegnet in zahlreichen Zusammenhängen und mit unterschiedlichen Bedeutungen. Für die vorliegende Studie ist es daher zunächst erforderlich, den Terminus im Kontext von Unterricht zu klären. Die Frage nach gutem Erklären einer Lehrkraft lenkt dann den Blick auf Konzepte von Unterrichtsqualität und fokussiert das Paradigma professioneller Kompetenz. In diesem Zusammenhang interessieren vor allem die Verfahren, mit Hilfe derer die Qualität von Unterricht beziehungsweise die (Erklär-)Kompetenz von Lehrkräften gemessen oder beurteilt werden kann. Auf dieser Basis werden Charakteristika unterrichtlichen Erklärens herausgearbeitet und in einer Definition gefasst, bevor daran anknüpfend Qualitätsmerkmale unterrichtlichen Erklärens entfaltet werden. Ausgehend von dieser in erster Linie fachübergreifenden Perspektive wird unterrichtliches Erklären abschließend im Kontext von Musikunterricht exploriert, wobei in diesem Zusammenhang ein Schwerpunkt auf der Vermittlung musiktheoretischer Unterrichtsinhalte liegt.

2.1 Zum Begriff Erklären

Der Begriff *Erklären*⁵ wird im Alltag vielfältig in Gebrauch genommen. Im Duden finden sich für das Verb drei allgemeinsprachliche Bedeutungsbeschreibungen, die teilweise noch in Unterbedeutungen aufgegliedert werden (Dudenredaktion, o.J.). In der ersten Kategorie steht *Erklären* in den Kontexten (1) *deutlich machen*, wenn beispielsweise ein Sachverhalt anhand eines Beispiels *erklärt* wird, (2) *begründen/ deuten*, wenn zum Beispiel *erklärt* wird, warum eine Person nicht anwesend sein kann, und (3) *eine Begründung in etwas finden*, wenn sich etwa ein Sachverhalt aus sich selbst heraus *erklärt* (Dudenredaktion, o.J.). Dabei wird ersichtlich, dass das Verstehen als Zielperspektive in allen drei Kontexten dieser Bedeutungsbeschreibung eine wichtige Rolle spielt. Die zweite Kategorie verwendet *Erklären* in Form von (1) *mitteilen/äußern*, wenn beispielsweise ein Rücktritt *erklärt* wird, oder als (2) *Verdeutlichung der eigenen Haltung*, indem man sich zum Beispiel einer Gruppe solidarisch *erklärt* (Dudenredaktion, o.J.). Die dritte Kategorie umfasst die Bedeutung im Kontext *bezeichnen/kennzeichnen*, wenn beispielsweise eine Person als Sieger *erklärt* wird (Dudenredaktion, o.J.). Für die vorliegende Arbeit ist die erste Kategorie relevant, da insbesondere das Verstehen im Gegensatz zu den beiden anderen Kategorien ein Charakteristikum darstellt, das für *unterrichtliches Erklären* als bedeutend erscheint.

Im wissenschaftlichen Sprachgebrauch differenziert sich *Erklären* in unterschiedlicher Weise aus. In linguistischer Fachliteratur wird zunächst festgestellt, dass das Konzept von *Erklären* „in einer diffusen Weise zwischen alltäglicher und theoretischer Sprache angesiedelt ist“ (Ehlich, 2009, S. 11). Aus fachsprachlicher Sicht wird *Erklären* als

5 Um Unterschiede zwischen Erklären und weiteren Sprachhandlungen deutlich zu machen, werden diese in Kapitel 2.1 an den entsprechenden Stellen kursiv gesetzt.

Sprachhandlung betrachtet, die in Fachkommunikation, in wissenschaftlichem Diskurs sowie in der Wissensvermittlung zwischen Expert*innen und Laien in Erscheinung treten kann (Hohenstein, 2009). Hinsichtlich einer sprachlichen Differenzierung wird angeregt, *Erklären* von semantisch naheverwandten und vermutlich deshalb häufig synonym genutzten Verben wie *Argumentieren*, *Begründen*, *Erzählen*, *Erläutern*, *Rechtfertigen* oder *Instruieren* möglichst trennscharf abzugrenzen (Ehlich, 2009; Hohenstein, 2009; Kotthoff, 2009).⁶ Allerdings wird im selben Atemzug festgestellt, dass dies bisher noch nicht zufriedenstellend eingelöst werden konnte; beispielsweise habe es den Anschein, dass einzelne Verben wie *Erklären* oder *Erläutern* zur gegenseitigen Definition herangezogen werden (Ehlich, 2009). Zu unterscheiden wären diese beiden Sprachhandlungen jedoch hinsichtlich des Bezugs zum Vorwissen der Adressat*innen – je weniger Vorwissen vorhanden ist, desto mehr muss unterstützt beziehungsweise *erklärt* werden, wohingegen beim *Erläutern* verstärkt auf bereits vorhandene Wissensstrukturen beziehungsweise Erklärungen aufgebaut werden kann (Kiel, 1999).

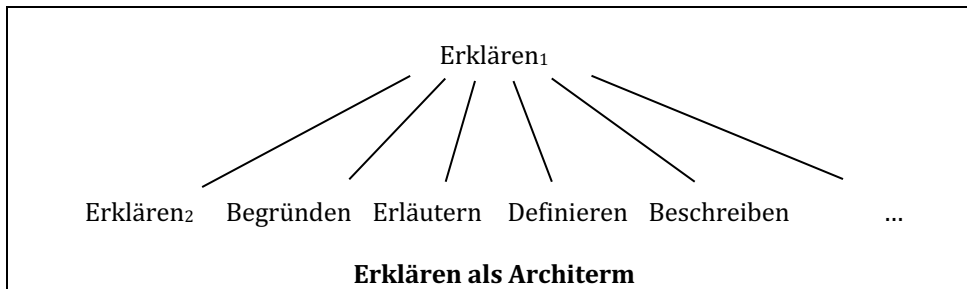
Nachvollziehbar gestaltet sich diesbezüglich auch eine Einteilung von Hohenstein (2009), die zwischen *Erklären*, *Erläutern*, *Begründen* und *Instruieren* unterscheidet und als Kriterium dafür die bei der Sprachhandlung jeweils verfolgte kommunikative Absicht anlegt: Während beispielsweise *Begründen* auf eine Umorganisation bestehender Wissensstrukturen und *Instruieren* auf die Befähigung zum Handlungsvollzug abzielt, verfolgt *Erklären* den Zweck, Wissen auf- beziehungsweise auszubauen (Hohenstein, 2009). An anderer Stelle führt Hohenstein (2006) aus, dass unter dem Zweck des Erklärens auch „komplexe, d. h. mehrfach zusammengesetzte Formen sprachlichen Handelns“ entstehen können, „bei denen Zusammenschlüsse von erklärenden mit beschreibenden, begründenden, erläuternden und illustrierenden (Teil-)Handlungen vorliegen“ (S. 113). Daher kann *Erklären* sowohl als einzelner illokutiver Akt als auch innerhalb eines Illokutionenverbunds⁷ mehrerer Sprachhandlungen auftreten (Hohenstein, 2006). In Anlehnung daran schlägt Neumeister (2011) ein Schema vor, das zwischen *Erklären*₁ und *Erklären*₂ unterscheidet. *Erklären*₁ versteht sich in einer eher alltäglich gebrauchten und weiter gefassten Bedeutung von Klarmachen und subsummiert als Oberbegriff die eben genannten Sprachhandlungen, zu denen auch *Erklären*₂ zu zählen ist, das als komplexe und in die Tiefe gehende Form der Wissensvermittlung angesehen wird. Im Sinne Hohensteins (2006) kann durch *Erklären*₂ Wissen auf- beziehungsweise ausgebaut werden, indem Verborgenes sichtbar gemacht und Funktionszusammenhänge dargestellt werden (Neumeister, 2011). Abbildung 2.1⁸ zeigt das strukturelle Verhältnis von *Erklären*₁ und *Erklären*₂.

6 Einige dieser Sprachhandlungen spielen im musikpädagogischen Diskurs eine Rolle, z. B. das *Argumentieren* (z. B. Ehninger et al., 2021; Rolle, 2014), das *Erzählen* (Cvetko, 2011; Oberhaus, 2015) oder das *Instruieren* (z. B. Frei et al., 2022; Stibi, 2018, 2019).

7 Zu den Begriffen *Illokution* und *Perlokution* siehe z. B. Klein, 2009, S. 27.

8 Die Nummerierung von Abbildungen und Tabellen folgt in dieser Arbeit dem Prinzip, das zunächst das Kapitel genannt wird und dann pro Kapitel alle Abbildungen nummeriert werden. In diesem Beispiel wäre Abbildung 2.1 also die erste Abbildung in Kapitel 2. Die Nummerierung startet bei jedem Kapitel wieder von vorne.

Abbildung 2.1 Erklären₁ und Erklären₂ (vgl. Neumeister, 2011, S. 95)



Die erziehungswissenschaftliche Perspektive rückt Erklären als didaktisches Handeln in den Blick. Dabei ist die Konzeptualisierung in erster Linie an die ursprüngliche Wortherkunft *deutlich* beziehungsweise *klar machen* angelehnt und betrachtet *Erklären* grundsätzlich als Komplement von Verstehen (Kiel, 1999; siehe ausführlich Kap. 2.3.1). Darüber hinaus wird, wie auch aus sozialwissenschaftlicher Sicht (Bartelborth, 2007), zwischen *Erklären* und *Erklärung* unterschieden: *Erklären* wird dabei als Prozess aufgefasst, wohingegen *Erklärung* das Produkt dieses Prozesses meint (Kiel, 1999). Da der Begriff *Erklärung* darüber hinaus auch im Kontext wissenschaftstheoretischer Erklärskemata verwendet wird, sollten wissenschaftstheoretische und unterrichtliche Konzepte von Erklären unterschieden werden (Schilcher et al., 2017). Beispielsweise skizzieren Hempel und Oppenheim (1948) in ihrem Modell, dass durch bestimmte Voraussetzungen (*Explanans*) wie Gesetzmäßigkeiten und Nebenbedingungen auf ein logisch ableitbares, beobachtbares und zu erklärendes Ergebnis oder Phänomen (*Explanandum*) geschlossen werden kann (Wagner & Wörn, 2011). Folgt eine Erklärung diesem Modell, kann sie als wissenschaftlich korrekt eingestuft werden.

Kiel (1999) erachtet es jedoch als problematisch, dass Modellierungen dieser Art ohne empirische Validierung auf Unterricht übertragen werden. Zwar lassen sich einige mathematische und naturwissenschaftliche Phänomene durch ein kausales Zurückführen erklären – das Schema von Hempel und Oppenheim (1948) wird beispielsweise im Kontext von Erklären im Physikunterricht rezipiert (Kulgemeyer & Tomczyszyn, 2015). Jedoch kritisieren Schilcher et al. (2017) die Anwendbarkeit eines solchen Schemas zur Bewertung von Erklären im Kontext von Unterricht gerade aus Sicht geisteswissenschaftlicher Fächer als verengend und sprechen diesem auch wegen der fehlenden Berücksichtigung von Qualitätsmerkmalen unterrichtlichen Erklärens eine fachübergreifende Gültigkeit ab. Bezüglich der Qualität einer unterrichtlichen Erklärung schlägt Kiel (1999) vor, „die Zufriedenheit der Gesprächsteilnehmer als wichtigstes Kriterium“ (S. 254) heranzuziehen – im unterrichtlichen Kontext also das Urteil von Schüler*innen. Auf beide zentralen Aspekte dieser Aussage geht die vorliegende Arbeit noch ausführlich ein: In Kapitel 2.2 werden Mess- sowie Beurteilungsverfahren, in Kapitel 2.3.2 Qualitätsmerkmale unterrichtlichen Erklärens herausgearbeitet.

Um *Erklären* als didaktisches Handeln im schulischen Kontext zu bestimmen, wird das Attribut *unterrichtlich* ergänzt (z. B. Neumeister, 2011; Schilcher et al., 2017). Damit

ist eine Abgrenzung zu anderen Bereichen gezogen, in denen *erklärt* wird, etwa zwischen IT-Fachkräften und Laien oder aber auch in der Kommunikation zwischen Ärzt*innen und Patient*innen (z. B. Hargie, 2013; Renkl et al., 2006). Da in diesem Zusammenhang auch die Möglichkeit besteht, dass nicht nur Lehrkräfte, sondern auch Schüler*innen ihren Klassenkamerad*innen etwas *erklären* können (z. B. Berkemeier, 2009; Kulgemeyer & Schecker, 2013), gelingt eine Fokussierung auf das Handeln von Lehrpersonen, indem dieses als *unterrichtliches Erklären einer Lehrkraft* bezeichnet wird (Lindl et al., 2019; siehe auch Schilcher et al., i. Vorb.). In diesem Zusammenhang wird auch der Unterschied zwischen *Erklären gegenüber einer weiteren Person* und *Sich-selbst-erklären* diskutiert: Wenn *Erklären* als interaktiver Kommunikationsprozess zwischen mindestens zwei Personen angesehen wird (Hohenstein, 2009; Leinhardt, 2001), wäre ein *Sich-selbst-erklären* von Schüler*innen unter dieser Definition nicht inkludiert (Wörn, 2014). Dass vor diesem Hintergrund andere Charakteristika beim *Sich-selbst-erklären* anzunehmen wären als beim *Erklären gegenüber einer weiteren Person*, wird an anderer Stelle erläutert (siehe Kap 2.3.1).

Erklären als unterrichtliches Handeln einer Lehrkraft wird in der erziehungswissenschaftlichen und fachdidaktischen Literatur unterschiedlich konzeptualisiert. Nach Find-eisen (2017) bestehe daher „Unklarheit darüber, ob das Erklären von Fachinhalten als Wissen, als Fähigkeit oder als Kompetenz verstanden wird“ (S. 100). In ihrer Arbeit selbst begründet sie ausführlich und nachvollziehbar, warum in der Abhandlung von Erklären als *Fähigkeit*⁹ gesprochen wird, wobei der Titel *Fachdidaktische Kompetenzen angehender Lehrkräfte* zunächst eine andere konzeptuelle Rahmung erwarten lässt (Find-eisen, 2017). Während auch andere Arbeiten von *Erklärfähigkeit* sprechen – wobei die genaue Benennung zwischen *Erklärfähigkeit* (Kulgemeyer & Schecker, 2013) und *Erklärerfähigkeit* (Kulgemeyer & Tomczyszyn, 2015) schwanken kann – wird es im Rahmen des kompetenztheoretischen Ansatzes häufig als *Erklärkompetenz* bezeichnet (z. B. Reusser & Pauli, 2003; Schopf & Zwischenbrugger, 2015; Spreckels, 2009b; Wörn, 2014; siehe dazu auch Kap. 2.2.2). Wörn (2014) kritisiert diesbezüglich jedoch, dass der Begriff *Erklärkompetenz* zwar beispielsweise bei Reusser und Pauli (2003) verwendet, jedoch dort nicht ausreichend definiert wird.

Die COACTIV-Studie (Kunter et al., 2011) untersucht das fachliche, fachdidaktische und pädagogisch-psychologische Wissen von Mathematiklehrkräften und spricht in Bezug auf Erklären auch konsequenterweise von *Erklärungswissen* beziehungsweise von „Wissen über multiple Repräsentations- und Erklärungsmöglichkeiten“ (Baumert & Kunter, 2011, S. 38). Sie rekurriert auf die Taxonomie über das Professionswissen von Lehrkräften nach Shulman (1986), in der das Wissen über Erklären, Darstellen und Verständlichmachen zu einer der wichtigsten fachdidaktischen Kompetenzen einer Lehrkraft gezählt wird: „(...) the most useful forms of representation (...), the most powerful analogies, illustrations, examples, explanations, and demonstrations – in a word, the ways of representing and formulating the subject that make it comprehensible to others“ (Shulman, 1986, S. 9–10). Dabei werden *the ways of representing* dem Bereich *pedagogical*

9 Um Unterschiede in den Konzeptualisierungen deutlich zu machen, werden auch diese Begriffe kursiv gesetzt.

content knowledge zugeordnet, was in der deutschsprachigen Fachliteratur mittlerweile nahezu einheitlich mit dem Begriff *fachdidaktisches Wissen* übersetzt wird (Krauss et al., 2011).

In der musikpädagogischen Forschungsliteratur findet das unterrichtliche Erklären bisher nur wenig Berücksichtigung. In der Studie FALKO-M (Puffer & Hofmann, 2017) wurde analog zu den anderen Fächern der interdisziplinären FALKO-Forschungsgruppe (Krauss et al., 2017) ein Professionswissenstest für die Bereiche Fachwissen und Fachdidaktisches Wissen von (angehenden) Musiklehrkräften nach dem Vorbild der Studie COACTIV (Kunter et al., 2011) und ebenfalls unter Rekurs auf die Typologie Shulmans (1986) entwickelt und validiert. Die musikspezifische Ausgestaltung dessen, was darin unter dem Wissensbereich *Erklären und Verständlichmachen* verortet ist, wird als „Wissen über musikbezogene Instruktionsstrategien“ (Puffer & Hofmann, 2017, S. 254) ausgewiesen. Der Terminus ist als Sammelbegriff im Sinne des Wortes *instruction* im angloamerikanischen Raum (z. B. Dorfman, 2013; Kunter et al., 2013) und schließt explizit Aspekte wie musikpraktisches Vormachen oder die Initiierung von Gestaltungsprozessen ein (Puffer & Hofmann, 2017).

Während diese fachspezifische Modellierung eine musikpraktische Dimension erkennen lässt, wird hingegen das Wort *Erklären* in diesem Zusammenhang ausgespart. Mit Blick auf die sprachwissenschaftliche Unterscheidung bleibt zudem unklar, in welchem Verhältnis Erklären und Instruieren – beides hier verstanden als unterrichtliches Handlungsmuster einer Musiklehrkraft – zueinanderstehen. Eine erste Modellierung guten Erklärens, die auch eine Unterscheidung zum Instruieren anhand des dabei verfolgten Zwecks (Hohenstein, 2009) berücksichtigt, findet sich an anderer Stelle (Frei et al., 2022) und wird in dieser Arbeit ausführlich in Kapitel 2.4.2 erläutert. Dabei wird herausgearbeitet, inwieweit die sprachwissenschaftliche Unterscheidung zwischen Erklären und Instruieren zu einem trennscharfen Bild über das Unterrichtshandeln einer Musiklehrkraft beitragen kann.

Neben den drei häufig zu findenden Begriffen *Erklärfähigkeit*, *Erklärkompetenz* und *Erklärungswissen* lassen sich auch noch weitere im Kontext unterrichtlichen Erklärens finden. Gindele und Voss (2017) verwenden in ihrer Analyse über Zusammenhänge zwischen pädagogisch-psychologischem Wissen und Indikatoren des beruflichen Erfolgs den Terminus *Erklärungsfähigkeit*, ohne das dahinterliegende Begriffsverständnis näher auszuführen. Da die Ergebnisse der Untersuchung auf Daten der COACTIV-Studie (Baumert & Kunter, 2011) beruhen, bleibt zudem unklar, in welchem Verhältnis *Erklärungsfähigkeit* und *Erklärungswissen* stehen. Eine weitere, etwas anders gelagerte, Perspektive suggeriert der Begriff *Erklärhandeln*, der in einem physikdidaktischen Forschungsvorhaben verwendet wird (Riese et al., 2015). Dieser richtet den Fokus auf die Performanz des Erklärens – wobei *Erklär-Handeln* streng genommen tautologisch ist – und impliziert im Vergleich zu *Erklärkompetenz*, *Erklärfähigkeit* oder *Erklärungsfähigkeit* weniger die Voraussetzungen, die für das Erklären notwendig sein könnten. Dieser bunte Strauß an Begrifflichkeiten kommt teilweise sogar innerhalb eines Forschungsberichts zum Tragen. In einem konkreten Fall stelle „*Erklärungsfähigkeit* [Hervorhebung vom Autor hinzugefügt] für alle am Unterricht Beteiligten eine wichtige *Kompetenz* [Hervorhebung vom Autor

hinzugefügt]“ (Spreckels, 2009b, S. 118) dar. Diesbezüglich besteht die Annahme, dass die Gleichsetzung beider Begriffe sowie die spätere Nennung von *Fertigkeiten*, die von Spreckels (2009b) im Kontext mündlichen Erklärens im Deutschunterricht erwähnt werden, mehr eine begriffliche Varianz denn einen trennscharfen Gebrauch zum Ziel hat. Unklar bleibt jedoch, in welcher Beziehung die drei Begriffe zueinanderstehen.

Zusammenfassend können zwei Leitlinien abgeleitet werden, an denen sich die terminologische Rahmung des Begriffs *Erklären* orientiert: *Erklären* verfolgt grundsätzlich das Ziel, Verstehensprozesse zu initiieren, und wird als didaktisches Handeln einer Lehrkraft im Unterricht angesehen (Kiel, 1999). Wirkmechanismen wissenschaftstheoretischer Erklärungen können für das Fach Musik aufgrund seiner geisteswissenschaftlichen und vor allem ästhetisch geprägten Ausrichtung im Gegensatz zu mathematisch-naturwissenschaftlichen Fächern (z. B. Kulgemeyer & Tomczyszyn, 2015) nicht ausreichend fruchtbar gemacht werden. Denn gerade die objektiven Gesetzmäßigkeiten eines deduktiv-nomologischen Modells, wie dem von Hempel und Oppenheim (1948), können subjektiv-ästhetische Erfahrungen der Schüler*innen entgegenstehen – diese zu ermöglichen, wird als ein zentrales Ziel von Musikunterricht ausgegeben (Rolle, 2010). In Anlehnung an Kiel (1999) hat das zur Folge, dass der Begriff *Erklärung* im musikpädagogischen Kontext daher nur verwendet werden sollte, wenn von dem Produkt eines Erklärprozesses gesprochen wird. Solche Produkte liegen in der vorliegenden Arbeit beispielsweise in Form der Videovignetten vor, die als Erklärungen benannt werden. Um den Fokus auf die Unterrichtstätigkeit und das Unterrichtshandeln von Musiklehrkräften zu richten, wird insbesondere der Begriff *Erklären* verwendet.

Hinsichtlich der verschiedenen Konzeptualisierungen von *Erklären* als didaktisches Handeln, in denen es als *Wissen*, *Fähigkeit*, *Kompetenz*, *Fertigkeit* oder auch *Handlung* bezeichnet wird, kann die in Anlehnung an Weinert (2001) vorgenommene Differenzierung von Klieme et al. (2003) ordnende Funktion übernehmen: Demnach umfassen *Kompetenzen* die neben der motivationalen, volitionalen und sozialen Bereitschaft zur Bewältigung von Problemsituationen – also *im Handeln* – bei Individuen verfügbaren beziehungsweise erlernbaren *Fähigkeiten und Fertigkeiten*. Das dazu erforderliche *Wissen* wird als Bestandteil von Kompetenz erachtet (Frey, 2014). Vor diesem Hintergrund und in Anlehnung an das Verständnis von Kompetenz als Kontinuum zwischen Disposition und Performanz nach dem Modell von Blömeke, Gustafsson et al. (2015) wird im Kontext dieses Forschungsvorhabens der Begriff *Erklärkompetenz* verwendet (siehe dazu ausführlich Kap. 2.2.2). Damit erfolgt die terminologische Rahmung von Erklären auf zwei Ebenen: Einerseits fokussiert die Darstellung im Folgenden, wenn nicht explizit anders geschrieben, das *unterrichtliche Erklären einer (Musik-)Lehrkraft*. Andererseits wird von Erklären als *Kompetenz* gesprochen.

2.2 Bildungswissenschaftlicher Diskurs

Innerhalb der Unterrichtsforschung spielt das unterrichtliche Erklären einer Lehrkraft eine bedeutende Rolle. Erklären wird nicht nur als „zentrale fachdidaktische Tätigkeit“

(Findeisen, 2017, S. 2) deklariert oder innerhalb des *core-practice*-Ansatzes zu den „Kernpraktiken“ (Fraefel & Scheidig, 2018, S. 344) im Unterricht gezählt. Nach Pauli (2015) gilt es als „Basiskompetenz von Lehrpersonen“ (S. 46), einen Sachverhalt gut erklären zu können. Schüler*innen benennen dies sogar als eine der wichtigsten Kompetenzen einer Lehrkraft (z. B. Wörn, 2014; Wragg & Wood, 1984). Die Frage nach der Qualität von Erklären sowie die Aussage, es gehöre zu den Kerntätigkeiten einer Lehrkraft, leiten zum allgemeinen Diskurs über Unterrichtsqualität und fokussieren die Rolle der Lehrperson, die als wichtiger Einflussfaktor für die Qualität von Unterricht gilt (z. B. Hattie et al., 2014; Kunter et al., 2011). Obwohl die bildungswissenschaftliche Forschung zu Unterrichtshandeln und Kompetenzen von Lehrkräften in den letzten Jahren deutlich zugenommen hat (Seidel, 2014), überrascht es, dass das Erklären bisher eine meist untergeordnete Rolle gespielt hat (Kiel, 1999, S. 15) und erst in den neuesten Handbüchern zur Lehrkräfteforschung an einzelnen Stellen thematisiert wird (siehe z. B. Cramer et al., 2020).

Im Folgenden wird in Kapitel 2.2.1 herausgearbeitet, welche Konzepte zur Beschreibung von Unterrichtsqualität vorliegen, welche Qualitätsmerkmale als relevant erachtet werden und an welchen Stellen dabei unterrichtliches Erklären Berücksichtigung findet. Der Darstellung forschungsmethodischer Verfahren und Instrumente, die bei der Untersuchung von Unterrichtsqualität zur Anwendung kommen, folgt ein Überblick über verschiedene Ansätze zur Forschung über Lehrkräfte. In einem zweiten Schritt (Kap. 2.2.2) wird das Paradigma professioneller Kompetenz erläutert, welches die zweite bildungswissenschaftliche Diskurslinie der vorliegenden Arbeit umfasst. Dabei wird entfaltet, wie Erklärkompetenz in der Forschungsliteratur konzeptualisiert und operationalisiert wird. Die Darstellung schließt mit forschungsmethodischen Verfahren und Instrumenten, mit Hilfe derer die Erklärkompetenz von Lehrkräften erhoben wird.

2.2.1 Unterrichtsqualität

Die bildungswissenschaftliche Forschung zu Unterrichtsqualität untersucht „diejenigen Aspekte von Unterricht, die das Verfolgen extern gesetzter Erziehungs- und Bildungsziele unterstützen“ (Praetorius & Nehring, 2020, S. 297). Dabei variieren innerhalb des Diskurses nicht nur die Zielvorstellungen von Unterricht, indem dieser beispielsweise *gut, erfolgreich, gelingend, effektiv, wirkungsvoll* oder auch *qualitätvoll* sein soll (z. B. Baumert & Kunter, 2011; Helmke, 2017; Klieme & Rakoczy, 2008; Kunter & Trautwein, 2013; Lipowsky, 2007; Meyer, 2004; Terhart, 2009). Sichtbar werden auch unterschiedliche Herangehensweisen und Zugänge: Während nach Meyer (2004) „normativ (also auf Grundlage einer Bildungstheorie) gesetzt“ (S. 12) sei, was guter Unterricht bedeute, richtet Lipowsky (2007) den Blick auf Merkmale guten Unterrichts, deren Wirksamkeit auf Basis der Lernerfolge von Schüler*innen bereits nachgewiesen wurde. Bevor Merkmale

guten Unterrichts vorgestellt werden, richtet sich der Blick zunächst auf grundlegende Konzepte von Unterrichtsqualität.¹⁰

Modellierung

Die Modellierungen von Erklärqualität kann sich am Diskurs um Unterrichtsqualität orientieren, der in den letzten zwei Dekaden eine Vielzahl an Modellen hervorbrachte (z. B. Brophy, 2000; Helmke, 2017; Kunter & Voss, 2011; Meyer, 2004; Praetorius, Rogh et al., 2020; Vieluf et al., 2020; Wisniewski, Zierer et al., 2020). Für den deutschsprachigen Diskurs ist dabei das Angebot-Nutzungs-Modell von Helmke (2017) prägend, das nicht nur als theoretischer Rahmen verschiedener, groß angelegter Studien aus dem Feld der Unterrichtsforschung dient (z. B. Blömeke et al., 2010a und 2010b; DESI-Konsortium, 2008; Wisniewski, Zierer et al., 2020), sondern auch bei empirischen musikpädagogischen Forschungsvorhaben als theoretischer Bezugspunkt herangezogen wird (z. B. Harnischmacher et al., 2017; Heß, 2018b). Helmke (2017) integriert in seinem fachübergreifend gültigen Modell sowohl Merkmale von Unterrichtsqualität als auch Zielkriterien des Unterrichts und zeigt jeweilige Wirkungsweisen zwischen einzelnen Komponenten auf. Die zentrale Achse des Modells setzt sich zusammen aus (1) dem Unterrichtsangebot, das von der Lehrperson gestaltet wird, (2) den Lernaktivitäten der Schüler*innen, die sich aus der individuellen Nutzung des Angebots ergeben, und (3) dem daraus resultierenden Ertrag in Form von Kompetenzerwerb der Lernenden (Helmke, 2017).

Das Interesse dieser Arbeit liegt auf dem guten Erklären von Musiklehrkräften und fokussiert damit in diesem Modell insbesondere die Komponente *Lehrperson*. Dabei spielen neben Persönlichkeitsmerkmalen wie Geduld sowie schul- und unterrichtsrelevanten Wertvorstellungen, Zielen und Orientierungen einer Lehrperson vor allem ihre „unterrichtsrelevante Expertise, das heißt die fachwissenschaftliche und fachdidaktische Expertise“¹¹ (Helmke, 2017, S. 76) eine entscheidende Rolle. Gutes Erklären einer Lehrkraft ist dabei insofern Teil von Unterrichtsqualität, als es fachübergreifend als fachdidaktische Kompetenz einer Lehrkraft modelliert wird (z. B. Krauss et al., 2017; Kunter et al., 2011). Dabei kann beim Erklären jedoch nicht von einem direkten Merkmal von Unterrichtsqualität gesprochen werden, da es sich beim Blick auf die Lehrperson grundsätzlich um „personale Merkmale handelt, die den Unterricht zwar beeinflussen, die aber nicht selbst als Aspekte der Unterrichtsqualität angesehen werden können“ (Helmke, 2017, S. 76).

Kunter und Trautwein (2013) modifizieren das Modell an verschiedenen Stellen, indem sie beim Unterricht zwischen Tiefenstrukturen wie dem Grad an kognitiver Aktivierung und Oberflächenstrukturen wie der Methode direkter Instruktion differenzieren (Kunter & Trautwein, 2013, S. 17). Decristan et al. (2020) merken diesbezüglich an, dass vor allem Tiefenmerkmale „besonderes Potenzial zur Unterstützung der Lernprozesse

10 Wie in der Einleitung erwähnt, werden die Begriffe *Guter Unterricht* und *Unterrichtsqualität* synonym verwendet.

11 Helmke (2017) verwendet die Begriffe *fachdidaktische Kompetenz* und *fachdidaktische Expertise* synonym und ist damit dem Paradigma professioneller Kompetenz zu verorten (Krauss, 2020). Eine ausführliche Darstellung des Paradigmas erfolgt in Kapitel 2.2.2.

von Schüler*innen haben“, jedoch Methoden wie direkte Instruktion¹² „zwar primär auf der Oberflächenebene verortet sind, aber durch ihre theoretischen Wirkmechanismen eng mit der Unterstützung kognitiver Prozesse verbunden sind“ (S. 111–112). Unterrichtliches Erklären einer Lehrkraft kann dabei in allen Phasen direkter Instruktion, insbesondere jedoch in darstellender Weise während der Darbietung sowie in korrigierender Funktion beim anschließenden Üben, eine bedeutende Rolle spielen (Kunter & Trautwein, 2013). Zudem hat es das Potenzial, kognitive Prozesse anzuregen und Schüler*innen kognitiv zu aktivieren beziehungsweise ihren Lernprozess zu unterstützen.¹³

Beide Modelle nehmen für das Ziel einer grundsätzlich fachübergreifenden Anwendbarkeit in Kauf, keine ausreichend fachspezifische Gültigkeit erlangen zu können (Helmke & Schrader, 2008). Kunter und Trautwein (2013) verweisen hinsichtlich des fachübergreifenden Bezugs auf den dafür notwendigen höheren Abstraktionsgrad und gebrauchen daher den Begriff „Rahmenmodell“, da dieser die Funktion einer „Vereinfachung der Wirklichkeit“ (S. 16) verdeutlicht. Ein fachspezifischer Rückgriff, wie im vorliegenden Fall aus der Perspektive von Musikunterricht, muss eine solche Limitation der Modelle berücksichtigen. Zu ergänzen wären hinsichtlich der Wirkungsweisen von Musikunterricht beispielsweise außerunterrichtliche Angebote wie Instrumental- oder Gesangsunterricht, da diese einen bedeutenden Einfluss auf die Lernaktivitäten einzelner Schüler*innen und damit auch auf den Unterricht beziehungsweise das Handeln der Lehrkraft einnehmen können (Jank, 2013). Das *integrierte Angebots-Nutzungs-Modell der Wirkweise des Unterrichts* (Vieluf et al., 2020) berücksichtigt diese Elemente und stellt eine weitere Modifizierung der bisher vorgestellten Modelle dar. Allerdings wird nach wie vor intensiv diskutiert, wie generisch oder fachspezifisch Modellierungen von Unterrichtsqualität angelegt sind (Praetorius & Gräsel, 2021).

Im Gegensatz zu diesen fachübergreifenden liegen auch mehrere fachspezifische Modelle von Unterrichtsqualität vor, die ausgewählte Merkmale guten Unterrichts fokussieren und bisher meist im Kontext von Mathematikunterricht entwickelt wurden (z. B. Klieme et al., 2001; Kunter & Voss, 2011). Die Modellierung zur Beschreibung von Unterrichtsqualität, die der COATIV-Studie zugrunde liegt, basiert auf den drei latenten Merkmalen kognitive Aktivierung, Klassenführung und konstruktive Unterstützung, welche anhand manifester Indikatoren wie dem Professionswissen von Lehrkräften oder der Qualität der ausgewählten Aufgaben operationalisiert werden (Kunter & Voss, 2011). Ursprünglich wurde das Modell dieser drei Basisdimensionen nach Klieme et al. (2001)

12 Zur Wirksamkeit direkter Instruktion siehe Hattie et al., 2014.

13 In diesem Kontext beobachten Schilcher et al. (2017), dass nach längerer Zeit einer vor allem konstruktivistisch orientierten Ausrichtung und dem Fokus auf Lernprozesse wieder vermehrt instruktionale Elemente des Lehrens ins Blickfeld von Unterrichtsforschung rücken. Wellenreuther (2013) argumentiert diesbezüglich aus einer pragmatischen Perspektive, dass es allein aufgrund der oft fehlenden Unterrichtszeit nicht möglich ist, dass sich Schüler*innen die fachlichen Inhalte ausschließlich selbst erarbeiten können. Der Aspekt mangelnder Unterrichtszeit ist dabei gerade für *kleine* Fächer wie Musik von entscheidender Bedeutung. Darüber hinaus plädieren Trautmann und Wischer (2016) für eine unterstützende und systematisierte Vermittlung durch Expert*innen, da im Kontext selbstgesteuerter Lernprozesse nicht immer eine ausreichende Verarbeitungstiefe von den Schüler*innen erreicht werden kann.

„mit dem Ziel eines aus wissenschaftstheoretischer Sicht anzustrebenden möglichst sparsamen Modells entwickelt“ (Praetorius, Rogh et al., 2020, S. 305). Auch in der Erweiterung um eine vierte Dimension, der *Strukturiertheit von Unterricht*, in der Pythagoras-Studie (Klieme & Rakoczy, 2008) werden damit jedoch insbesondere die Prozessqualität von Unterricht fokussiert und weitere Faktoren wie Schulform oder Zusammensetzung der Klasse ausgeblendet (z. B. Klieme et al., 2001; Kunter & Voss, 2011).

In der musikpädagogischen Studie ProBiNi adaptierten Haas et al. (2019) das Modell der drei Basisdimensionen in Anlehnung an Fauth et al. (2014), um die Frage zu beantworten, wie sich die Unterrichtsqualität zwischen Musikklassen¹⁴ und normalem Musikunterricht aus Sicht von Schüler*innen unterscheidet. Darüber hinaus nehmen auch weitere musikpädagogische Forschungsvorhaben auf das Modell der drei Basisdimensionen Bezug (z. B. Gebauer, 2016; Heß, 2018b; Niessen, 2010).

Dass jedoch alleine diese drei Merkmale, die vorwiegend aus der Perspektive des Mathematikunterrichts entwickelt wurden, die Qualität von Musikunterricht umfassend und valide beschreiben können, wird nicht nur innerhalb des musikpädagogischen Diskurses angezweifelt (z. B. Gebauer, 2016; Kranefeld, 2021; Puffer, 2021; Puffer & Hofmann, 2016) – auf die Notwendigkeit einer fachspezifischen Adaption beziehungsweise Modifikation wird auch im bildungswissenschaftlichen Diskurs hingewiesen (Klieme, 2006; Praetorius, Herrmann et al., 2020). Zudem etabliert sich in der bildungswissenschaftlichen Forschungsliteratur eine weitere Diskurslinie: Verschiedene Befunde aus der Unterrichtsforschung weisen darauf hin, dass gerade die Sparsamkeit des Modells der drei Basisdimensionen zu blinden Flecken bei der Beschreibung von Unterrichtsqualität führen kann (Praetorius, Rogh et al., 2020; siehe dazu auch Praetorius et al., 2018). Zwei aktuelle Modellierungen, die diese Mängel aufgreifen und neben einer generell fachübergreifenden Ausrichtung Ergänzungen der drei Basisdimensionen vorschlagen (Praetorius, Rogh et al., 2020; Wisniewski, Zierer et al., 2020), werden abschließend auch hinsichtlich einer umfassenderen Beschreibung von Unterrichtsqualität im Fach Musik vorgestellt – nicht zuletzt wird das Potenzial des Syntheseframework (Praetorius, Rogh et al., 2020) hinsichtlich einer musikpädagogischen Adaption erörtert (Kranefeld, 2021).

Basierend auf dem 7C-Modell des MET-Projekts (Measures of Effective Teaching, Ferguson & Danielson, 2014) validierten Wisniewski, Zierer et al. (2020) das so genannte *teaCh-Modell* anhand einer umfassenden Stichprobe ($N > 15.0000$ Schüler*innen). Es umfasst sieben Dimensionen: *Control* (z. B. Klassenführung), *Captivation* (z. B. Methodenvielfalt), *Conferment* (z. B. Feedback) und *Care* (z. B. emotionale Unterstützung und Ermutigung der Schüler*innen), *Challenge* (herausfordernde, kognitiv aktivierende Aufgaben), *Clarity* (Kohärenz und Struktur) sowie *Consolidation* (z. B. Sicherung durch Üben). Die Modellierung orientiert sich jedoch nicht nur an den drei Basisdimensionen, sondern zeigt sich auch gegenüber dem Angebot-Nutzungs-Modell theoretisch anschlussfähig (siehe Schilcher et al., i. Vorb.).

14 Musikklassen beziehungsweise Profilklassen (siehe dazu Heß, 2018a) sind Oberbegriffe für Streicher-, Bläser-, Chor- oder Bandklassen. In der zitierten Studie bleibt unklar, welche genauen Formen von Musikklassen an der Untersuchung teilgenommen haben.

Eine ebenfalls siebendimensionale Struktur umfasst die Modellierung von Praetorius und Charalambous (2018), die eine Synthese aus international vorliegenden Beobachtungsframeworks zur Unterrichtsqualität im Mathematikunterricht darstellt und sowohl mathematikspezifische, generische als auch hybride generisch-mathematikspezifische Frameworks beinhaltet – integriert ist beispielsweise auch das Modell der drei Basisdimensionen (Klieme et al., 2001) als eines der generischen Frameworks. In seiner deutschen Übersetzung erhält das Syntheseframework – im Folgenden Grundlage der Darstellung – eine generische Ausrichtung und umfasst ebenfalls sieben Dimensionen: (1) Angemessene Inhaltsauswahl und -thematisierung, (2) Kognitive Aktivierung, (3) Konsolidierung des Gelernten, (4) Formatives Assessment, (5) Unterstützung aller Schüler*innen durch Differenzierung und Adaptivität, (6) Sozio-emotionale Unterstützung und (7) Klassenführung (Praetorius, Rogh et al., 2020). Während ein entscheidender Unterschied zum teaCh-Modell die bisher noch ausstehende empirische Validierung darstellt, haben beide Modellierungen auf Ebene der Hauptdimensionen beispielsweise die Integration des Aspekts *Konsolidierung* beziehungsweise *Consolidation* gemein (siehe Praetorius, Rogh et al., 2020; Wisniewski, Zierer et al., 2020).

Zur musikpädagogischen Diskussion um Unterrichtsqualität können beide Modelle gerade wegen der besonderen Relevanz des Übens hinsichtlich musikbezogener Lernprozesse (Platz & Lehmann, 2018) einen wichtigen Beitrag liefern. Bezüglich des Erklärens konnte Kulgemeyer (2019) zeigen, dass eine anschließende Anwendung der Inhalte wirksam ist. Allerdings weisen beide Modellierungen auf Ebene der Hauptdimensionen auch unterschiedliche Schwerpunktsetzungen auf, die aus musikpädagogischer Perspektive relevant sind: Wisniewski, Zierer et al. (2020) fokussieren beispielsweise den Aspekt der Motivation von Schüler*innen (siehe für den musikpädagogischen Diskurs z. B. Harnischmacher, 2018), wohingegen Praetorius, Rogh et al. (2020) den Umgang mit Heterogenität stärker in den Blick nehmen, indem sie die Unterstützung aller Schüler*innen hinsichtlich eines differenzierten Lernangebots als eigenständige Dimension abbilden (siehe für den musikpädagogischen Diskurs z. B. Jank, 2013; Niessen, 2013).

In den bisherigen Ausführungen über Modelle zur Beschreibung von Unterrichtsqualität wurden verschiedene Anknüpfungspunkte für die Modellierung von Erklärqualität aufgezeigt. Gleichzeitig wurde die Relevanz verschiedener Merkmale deutlich, weshalb diese im Folgenden ausführlich erläutert werden. Dabei orientiert sich die Darstellung trotz fehlender empirischer Validierung an den Dimensionen des Syntheseframeworks (Praetorius, Rogh et al., 2020), da einerseits Prozessmerkmale von Unterrichtsqualität fokussiert und andererseits gerade die eben skizzierten Aspekte des Übens sowie des Umgangs mit Heterogenität für den musikpädagogischen Diskurs um Unterrichtsqualität bedeutsam scheinen. Darüber hinaus wurden bereits erste Aspekte der von Praetorius, Rogh et al. (2020) geforderten fachspezifischen Ausdifferenzierung aus musikdidaktischer Perspektive diskutiert (Kranefeld, 2021).

Zusammenfassend können für die Modellierung von Erklärqualität folgende Aspekte aus dem Diskurs um Unterrichtsqualität abgeleitet werden: Der Blick richtet sich primär auf die Angebotsseite von Unterricht, weshalb der Lehrkraft eine wichtige Rolle zukommt (z. B. Vieluf et al., 2020). Die Modellierung erfolgt im Spannungsfeld zwischen

Generik und Spezifik (Praetorius & Gräsel, 2021). Bei einer Adaption auf den Musikunterricht werden daher eigene fachspezifische Dimensionen diskutiert (Kranefeld, 2021). Zudem speist sich Unterrichtsqualität aus verschiedenen Merkmalen, die zur Qualität beitragen. Im Folgenden wird daher nicht nur beleuchtet, inwieweit das unterrichtliche Erklären einer Lehrperson mit Merkmalen guten Unterrichts zusammenhängt, sondern auch herausgearbeitet, welche dieser Merkmale den musikpädagogischen Diskurs von Unterrichtsqualität prägen.

Merkmale

Nach Helmke (2017) ist guter Unterricht von Merkmalen wie Klarheit und Strukturiertheit, Phasen der Konsolidierung und Sicherung oder dem Potenzial an kognitiver Aktivierung abhängig. Während sich diese Aspekte „direkt auf die Förderung der Informationsverarbeitung“ und damit auch den Lernerfolg von Schüler*innen beziehen, sind weitere Faktoren, wie Motivierung der Schüler*innen oder ein lernförderliches Klima im Unterricht, „primär auf die Förderung der Lernbereitschaft und indirekt auf den Lernerfolg“ (Helmke, 2017, S. 169) ausgerichtet. Jedoch ist allen genannten Merkmalen gemein, dass sie sich direkt auf die Prozessqualität von Unterricht beziehen. Damit unterscheiden sie sich von den weiter vorne erwähnten personalen Merkmalen der Lehrperson (z. B. die fachwissenschaftliche oder fachdidaktische Kompetenz), die den Unterricht beziehungsweise dessen Qualität nur indirekt beeinflussen (Helmke, 2017). Wie bereits dargestellt, dominieren dabei insbesondere die Merkmale Kognitive Aktivierung, Klassenführung und konstruktive Unterstützung¹⁵, die immer wieder als stabile Faktoren für die Wirksamkeit von Unterrichtsqualität nachgewiesen wurden (Praetorius et al., 2018; Wisniewski, Zierer et al., 2020) und den deutschsprachigen Diskurs der letzten Jahre prägen (Praetorius, Rogh et al., 2020).

Trotz dieser wichtigen Impulse für die Unterrichtsforschung wurde das Modell der drei Basisdimensionen gerade wegen seiner Sparsamkeit modifiziert und erweitert (Praetorius, Rogh et al., 2020). Mit dem Ziel, eine detailliertere Aussagefähigkeit über die Qualität von Unterricht zu erhalten, legen Kleickmann et al. (2020) eine Differenzierung der Dimension konstruktiver Unterstützung nahe. Unter Berücksichtigung der Merkmale kognitive Aktivierung und Klassenführung unterscheiden die Autor*innen zwischen emotionaler sowie kognitiver Unterstützung und können die prädiktive Validität dieses vierstufigen Modells nachweisen (Kleickmann et al., 2020). Emotionale Unterstützung deckt sich dabei weitestgehend mit dem bisherigen Verständnis konstruktiver Unterstützung und kann sich beispielsweise in einem wertschätzenden Umgang zwischen den am Unterricht Beteiligten äußern; im Gegensatz dazu spiegeln Reduktion von Komplexität oder inhaltliche Kohärenz im Unterricht Formen kognitiver Unterstützung wider

15 Kleickmann et al. (2020) weisen auf die unterschiedliche Bezeichnung dieser Basisdimension (z. B. unterstützendes Unterrichtsklima, siehe Lipowsky, 2009; konstruktive Unterstützung, siehe Kunter & Voss, 2011; lernförderliches Klima: z. B. Helmke, 2017) hin. Die vorliegende Arbeit folgt der Bezeichnung konstruktive Unterstützung, wie sie bei Kunter und Voss (2011) oder Praetorius, Herrmann et al. (2020) verwendet wird.

(Kleickmann et al., 2020) – Aspekte, die auch im Kontext unterrichtlichen Erklärens als Merkmale diskutiert werden (z. B. Findeisen, 2017; siehe ausführlich Kap. 2.3). Die begriffliche Nähe von kognitiver Aktivierung und kognitiver Unterstützung wird dabei mit Verweis auf das CLASS-Modell, in dem beide Dimensionen unter dem Begriff *instructional support* zusammengefasst sind, auch inhaltlich untermauert (Pianta & Hamre, 2009).

Das Syntheseframework von Praetorius, Rogh et al. (2020; siehe auch Praetorius & Charalambous, 2018) beschreibt Unterrichtsqualität anhand der oben aufgezählten sieben Dimensionen, wobei die Dimensionen *sozio-emotionale Unterstützung*, *Klassenführung* und *kognitive Aktivierung* dem Konzept der Basisdimensionen sehr ähnlich ist und diese „vollständig mit allen Subdimensionen abgedeckt sind“ (Praetorius, Rogh et al., 2020, S. 310). Die weiteren vier Hauptdimensionen des Syntheseframeworks richten jedoch einen zusätzlichen Fokus auf Merkmale, die bereits aus anderen Modellierungen bekannt sind: (1) Die Auswahl und Thematisierung von Inhalten inkludiert Aspekte wie fachliche Korrektheit, Konzepte wie Alltags- und Lebensweltbezug sowie Strukturierung von Inhalten (Praetorius, Rogh et al., 2020; siehe dazu z. B. Helmke, 2017; Klieme & Rakoczy, 2008). Auch die (2) Unterstützung aller Schüler*innen beinhaltet Aspekte von Adaptivität und Differenzierung bei den ausgewählten Inhalten des Unterrichts oder Aufgabenstellungen für die Lernenden und fokussiert damit den Umgang mit Heterogenität (Praetorius, Rogh et al., 2020; siehe z. B. Helmke, 2017). Dass Schüler*innen Gelegenheiten innerhalb des Lernprozesses erhalten, um Gelerntes durch viele Wiederholungen oder Variation von Aufgaben abzurufen, wird unter dem Aspekt des (3) Übens subsummiert (Praetorius, Rogh et al., 2020; siehe z. B. Helmke, 2017; Kunter & Trautwein, 2013). Die (4) Beurteilung von Schüler*innenleistungen in Form qualitativ hochwertigen Feedbacks, regelmäßiger Überprüfung des Verständnisses der Schüler*innen oder eines klaren Bezugs auf die als Ziel des Unterrichts zu erlernenden Kompetenzen runden das Framework ab (Praetorius, Rogh et al., 2020; siehe z. B. Hattie & Timperley, 2007; Helmke, 2017; Lipowsky, 2007; Wisniewski, Zierer & Hattie, 2020).

Um herauszuarbeiten, inwieweit Erklären im Unterricht beziehungsweise die Erklärkompetenz einer Lehrkraft als deren personales Merkmal auf Prozessmerkmale guten Unterrichts einwirken kann, orientiert sich die folgende Darstellung an den sieben Dimensionen des Syntheseframeworks nach Praetorius, Rogh et al. (2020) und fokussiert zunächst die drei Basisdimensionen (z. B. Lehner, 2018). Nach Pauli (2015) zeichnet sich gutes Erklären vor allem dadurch aus, kognitive Aufbauprozesse der Schüler*innen optimal zu unterstützen. In diesem Zusammenhang dient die Differenzierung von Kleickmann et al. (2020), die den Unterschied zwischen kognitiver und sozio-emotionaler Unterstützung in einem vierfaktoriellen Modell nachweisen konnten, als Schärfung der Basisdimension *konstruktive Unterstützung* hinsichtlich einer in diesem Kontext verstandenen kognitiven Unterstützung. In der Pythagoras-Studie wurde der Grad an *Unterstützung durch den Lehrer* unter anderem durch das Item „Unser Lehrer/Unsere Lehrerin erklärt etwas so lange, bis wir es verstehen“ (Klieme & Rakoczy, 2008, S. 231) operationalisiert, weshalb auch hier zwar einerseits die kognitive Unterstützung fokussiert wird, andererseits jedoch auch Elemente der sozio-emotionalen Unterstützung durchscheinen, indem der Lehrkraft implizit auch eine gewisse Geduld („erklärt etwas *so lange* [Hervorhebung vom Autor

hinzugefügt], bis ...“) attestiert wird. Vor diesem Hintergrund schreibt Brown (2006) dem Erklären, wenn es zu Verstehen führt, auch die Funktion zu, Ängste der Adressat*innen abzubauen.

Die Verbindung zwischen gutem Erklären und kognitiv aktivierendem Unterricht wird nicht nur durch die kognitiv geprägten Konzeptualisierungen von Erklären in verschiedenen Professionswissenstests von Lehrkräften deutlich (z. B. Krauss et al., 2017; Kunter et al., 2011). Auch Praetorius, Rogh et al. (2020) deklarieren unter anderem Erklären, Begründen, Argumentieren oder das Entwickeln von Ideen und Hypothesen als Handlungen, die Schüler*innen kognitiv aktivieren können. Auch die weitgehend geteilte Vorstellung, die Verstehen als grundlegendes Ziel unterrichtlichen Erklärens darstellt (z. B. Kiel, 1999; Findeisen, 2017), impliziert Erklären als kognitiv aktivierendes Moment. Demnach kann gutes Erklären als Teil eines kognitiv aktivierenden Unterrichts kognitive Verarbeitungsprozesse von Schüler*innen nicht nur unterstützen, sondern überhaupt erst aktivieren (Pauli, 2015).

Auch auf die dritte Basisdimension von Unterrichtsqualität, Klassenführung, kann gutes Erklären positiv Einfluss nehmen. Im Fach Mathematik konnte bereits gezeigt werden, dass ein Mangel an fachlicher Korrektheit zu Unterrichtsstörungen führen kann (Brunner, 2017). Zwar dürfte es nach Helmke (2017) unbestritten sein, dass sich fehlende fachliche Korrektheit negativ auf die Qualität des Unterrichts auswirkt, und auch hinsichtlich Qualitätsmerkmalen guten Erklärens wird die Notwendigkeit dieser Voraussetzung als trivial erachtet (Schilcher et al., 2017, S. 443), allerdings ist Erklären als eine der Kerntätigkeiten im Unterricht ebenfalls der Prämisse fachlicher Korrektheit verpflichtet (Wagner & Wörn, 2011; siehe Kap. 2.3.2). Daher wird auch ein Zusammenhang zwischen gutem Erklären und Klassenführung angenommen.

Die Inhalte der weiteren vier Dimensionen des Syntheseframework (Praetorius, Rogh et al., 2020) lassen sich ebenfalls direkt oder indirekt mit Erklären in Beziehung setzen. Die Dimension *Auswahl und Thematisierung von Inhalten und Fachmethoden* beinhaltet Aspekte wie die Vermeidung inhaltlicher Fehler beziehungsweise den Umgang mit präziser Fachsprache sowie die Strukturierung von Inhalten beziehungsweise Lernprozessen (Praetorius, Rogh et al., 2020). Erklären kann einerseits dazu beitragen, Verbindungen zwischen neuen sowie bereits gelernten Lerninhalten innerhalb einer Unterrichtsstunde zu verdeutlichen oder Bezüge zu vorherigen Stunden herzustellen (Wellenreuther, 2013), und andererseits Fachbegriffe sowie deren Bedeutung einführen (Findeisen, 2017). Aspekte der Dimension *formatives Assessment* sind beispielsweise die „regelmäßige Überprüfung des Verständnisses der Schüler*innen“ sowie ein „qualitativ hochwertiges Feedback“ (Praetorius, Rogh et al., 2020, S. 312). Eine Lehrkraft kann das Verständnis der Lernenden entweder durch gezieltes Nachfragen überprüfen oder anhand von Beiträgen, Rückmeldungen oder Fragen von Schüler*innen etwaige Fehlvorstellungen erkennen (Wagner & Wörn, 2011). Erklären kann dabei Teil des Feedbackprozesses sein und fehlende Informationen nachliefern beziehungsweise Fehlvorstellungen richtigstellen.

Schüler*innen beim Lernen und Anwenden neu erworbenen Wissens beziehungsweise neu erworbener Fähigkeiten zu unterstützen, stellt nach Praetorius, Rogh et al. (2020) ein weiteres Merkmal von Unterrichtsqualität dar. Unterrichtsliches Erklären einer

Lehrkraft kann dieser konstruktivistischen Phase des Übens vorausgehen (Kiel, 1999; Wagner & Wörn, 2011), sollte währenddessen jedoch eine untergeordnete Rolle spielen. Geht man umgekehrt von der Perspektive des Erklärens aus, konnten Wittwer und Renkl (2008) nachweisen, dass das Erklären einer Lehrkraft gerade dann besonders wirksam ist, wenn gute Lernaufgaben folgen, anhand derer das Erklärte von den Schüler*innen angewendet und demzufolge besser verstanden werden kann. Auch bei der Dimension *Unterstützung des Lernens aller Schüler*innen* tritt das Erklären in den Hintergrund und spielt keine explizite Rolle (Praetorius, Rogh et al., 2020). Das Erklären sollte sich zwar am Wissen der Adressat*innen orientieren (z. B. Findeisen, 2017), allerdings ist anzunehmen, dass eine Erklärung in der Regel nicht für alle Schüler*innen gleichermaßen verständlich sein kann.

Anhand dieser Gegenüberstellung von Prozessmerkmalen guten Unterrichts und dem Einsatz von Erklären kann weiterhin davon ausgegangen werden, dass sich Merkmale von Erklärqualität und Unterrichtsqualität überschneiden und beide Bereiche eng miteinander verbunden sind. Im Folgenden steht der musikpädagogische Diskurs um Merkmale guten Musikunterrichts im Fokus, während Anwendungsbereiche und Bedarfe am Erklären im Musikunterricht in Kapitel 2.4 herausgearbeitet werden.

Merkmale guten Musikunterrichts

Kranefeld (2021) gibt einen Überblick zum Stand der Diskussion um Unterrichtsqualität in der Musikdidaktik und resümiert, dass bisher noch kein umfassendes, konsensuales Modell vorliegt. Die Autorin führt dies insbesondere auf die „Pluralität der Konzeptionen und Ansätze in der Musikdidaktik“ (Kranefeld, 2021, S. 221) zurück, gibt aber zu erkennen, dass einzelne Aspekte in theoretischen musikpädagogischen Beiträgen mitverhandelt werden. Obwohl unter anderem Basisdimensionen guten Unterrichts, wie kognitive Aktivierung oder Klassenführung, bei der Frage nach Merkmalen guten Musikunterrichts aufgegriffen werden (z. B. Gebauer, 2013; Haas et al., 2019; Niessen, 2010; Puffer & Hofmann, 2017), orientiert sich der musikdidaktische Diskurs nicht primär an diesen bildungswissenschaftlichen Dimensionen.

In Anlehnung an die Klassifikation von Unterrichtsqualität nach Berliner (2005) ist die musikdidaktische Diskussion auf einer normativen Ebene zu verorten (Kranefeld, 2021); Erkenntnisse über die Wirksamkeit von Musikunterricht, gemessen am Lernerfolg von Schüler*innen, liegen bisher nicht vor (Puffer & Hofmann, 2017; siehe auch Heß, 2018b). Begründet wird dies beispielsweise damit, dass das Handeln und die Kompetenz von Musiklehrkräften bislang nicht im Zentrum des musikpädagogischen Forschungsdiskurses standen (Lehmann-Wermser & Krause-Benz, 2013). Außerdem „existieren bislang keine Analyseinstrumente zur Erfassung und Beschreibung inhaltlich-methodischer Merkmale von Musikunterricht“ (Heß, 2018b, S. 59). Auf Basis einer eher heuristischen Herangehensweise trägt Krupp (2021) anhand verschiedener Perspektiven von Expert*innen Aspekte eines wirksamen Musikunterrichts zusammen, legt dabei jedoch bewusst „kein bestimmtes Verständnis von Wirksamkeit“ (S. 8) zugrunde. Die in den Interviews

verhandelten Inhalte sind demnach im Sinne Berliners (2005) als Merkmale guten Unterrichts einzuordnen.

Einige der von Kranefeld (2021) benannten theoretischen Beiträge, in denen Merkmale guten Musikunterrichts (mit-)verhandelt werden (z. B. Dreßler, 2016; Fuchs & Brunner, 2015; Hasselhorn & Lehmann, 2014; Jank, 2013; Kaiser, 2010; Kraemer, 2007; Niessen, 2010; Rolle, 1999; Wallbaum, 2010), greifen Konsenslinien auf, die Puffer und Hofmann (2017) zur Gestaltung von gutem Musikunterricht identifiziert haben. Dieser soll Schüler*innen

- (a) musikalisch-ästhetische Erfahrungen ermöglichen,
- (b) zu einer verständigen Musikpraxis führen, indem er sich vor allem an außerschulischen musikbezogenen Erfahrungen der Schüler*innen orientiert, und
- (c) durch gemeinsame musikalische Praxis (Singen, Musizieren, Improvisieren, Komponieren und Musikhören) einen Beitrag zur musikalischen Bildung leisten (Puffer & Hofmann, 2017, S. 249; siehe auch Frei et al., 2022).

Auf Basis dessen werden im Folgenden weitere Merkmale zusammengetragen, die teilweise mit diesen drei Konsenslinien eng in Verbindung stehen. Abschließend wird reflektiert, was die einzelnen Merkmale für Erklären im Musikunterricht bedeuten können.

Aus einer praxeologischen Perspektive sollen in einem guten Musikunterricht möglichst alle Ausdrucksformen von Musik einbezogen werden und „musikalische Vielfalt als Normalfall“ (Krupp, 2021, S. 172; siehe dazu auch Niessen, 2010) gelten. Damit verknüpft werden weitere Forderungen an guten Musikunterricht: Neben der Förderung musikpraktischer Fähigkeiten (z. B. Benz, 2021; Fuchs & Brunner, 2015) soll guter Musikunterricht auch zum Nachdenken und Sprechen über Musik anregen (Niessen, 2010, S. 63). Nach Krupp (2021) sollen „ästhetische Erfahrungen und der Aufbau musikbezogenen Wissens und musikalischer Fähigkeiten [...] stets miteinander vernetzt werden“, mit Blick auf die Begründung und Formulierung ästhetischer Urteile seien demnach „sprachliche Mittel, aber auch bereits durchlaufene Erkenntnisprozesse erforderlich“ (S. 169). Die Bedeutung solcher inhaltlichen Aspekte sowie deren sprachliche Realisierung wird in verschiedenen Fachdidaktiken mit dem Begriff der Sachfachliteralität (Krupp, 2021) verdeutlicht – Tenorth (2004) verwendet hierfür den Terminus „kulturelle Literalität“ (S. 654).¹⁶ Unklar ist in diesem Zusammenhang die Bedeutung musiktheoretischer Wissensinhalte sowie die Frage nach Formen ihrer Vermittlung. Mit Blick auf die innerhalb des musikpädagogischen Diskurses geäußerte Kritik über Sinn und Zweck der Vermittlung resümiert Weidner (2015): „Oft wird dabei die Auffassung deutlich, ‚schlechten‘ Musikunterricht habe man vor allem bestimmten Formen musiktheoretischen Unterrichts zu verdanken, bei dem die ‚musikalische Praxis‘ zu kurz komme“ (S. 86).

Das in den Bildungswissenschaften mitunter zentrale Qualitätsmerkmal eines kognitiv aktivierenden Unterrichts wird mit Blick auf den Musikunterricht an verschiedenen Stellen der musikdidaktischen Forschung als nicht hinreichend deklariert (z. B. Gebauer, 2013; Puffer & Hofmann, 2016 und 2017). Mit Blick auf die Frage, wie ein guter

16 Dieser Diskurs wird aus mit Blick auf das Erklären musiktheoretischer Inhalte in Kapitel 2.4.3 vertieft.

Musikunterricht Schüler*innen ästhetische Erfahrungen ermöglichen kann, schlagen Puffer und Hofmann (2017) vor, dem Merkmal kognitiver Aktivierung auch das einer „ästhetischen Aktivierung“ (S. 250) an die Seite zu stellen.¹⁷ Obwohl eine konkrete inhaltliche Ausgestaltung noch aussteht, sieht Kranefeld (2021) darin eine potenzielle Kandidatin für ein grundsätzlich fachspezifisches Merkmal guten Musikunterrichts. Sie folgt damit der Aufforderung von Praetorius, Rogh et al. (2020), den fachübergreifend ausgerichteten Syntheseframework bei Bedarf um fachspezifische Merkmale zu ergänzen.

Vor diesem Hintergrund bringt Kranefeld (2021) mit der Unterstützung des Übens ein Merkmal in den Diskurs um guten Musikunterricht ein, welches aus musikpsychologischer und instrumentalpädagogischer Perspektive bereits eine bedeutende Rolle spielt (Platz & Lehmann, 2018). Guter Musikunterricht sollte dabei nicht nur ausreichend Raum für das Üben bieten, die Musiklehrkraft übernimmt dabei auch unterstützende Funktion, indem sie den Übeprozess einleitet, begleitet und dazu Rückmeldungen gibt (Fuchs, 2015; Krupp, 2021). Zudem erfordert Singen, instrumentales Musizieren und Bewegen zur Musik einen Unterricht, der auch auf sensomotorischer Ebene aktivierend sein kann (Kranefeld, 2021; siehe z. B. auch Krause-Benz, 2016). Neben diesen kognitiven, ästhetischen und sensomotorischen Formen der Unterstützung, werden auch sozio-emotionale Aspekte auf der Beziehungsebene zwischen Schüler*innen und der Lehrkraft, insbesondere in Situationen des gemeinsamen Musizierens genannt (Kranefeld, 2021; Krupp, 2021; zum Unterschied zwischen kognitiver und sozio-emotionaler Unterstützung siehe Kleickmann et al., 2020).

Neben diesen Merkmalen, die im Kontext des Syntheseframeworks von Praetorius, Rogh et al. (2020) verhandelt werden, fokussiert Kranefeld (2021) im Diskurs um Merkmale guten Musikunterrichts außerdem eine materiale Dimension von Unterricht – ein Aspekt, der im musikdidaktischen Diskurs zunehmend ins Zentrum des Forschungsinteresses gerückt wird. Die Autorin verweist dabei auf den Vorschlag, ein „didaktisches Tetraeder“ (Hußmann et al., 2018, S. 12) zur Modellierung der Unterrichtsgestaltung zu verwenden. Dieses greift die Idee des didaktischen Dreiecks auf, das aus dem Gegenstand, den Schüler*innen und der Lehrkraft besteht und ergänzt es um einen vierten materialen Aspekt (Hußmann et al., 2018). Mit Blick auf das Erklären verweist Godau (2018) auf die Klaviatur als häufig im Musikunterricht eingesetztes Material, das erklärende Funktion einnehmen kann.

Zusammenfassend erhält der auf bildungswissenschaftlicher Ebene breit geführte Diskurs um Unterrichtsqualität verstärkt Einzug in die musikdidaktische Unterrichtsfor-schung. Da bisher jedoch sowohl eine einheitliche Modellierung von Unterrichtsqualität (Kranefeld, 2021) als auch Erkenntnisse darüber fehlen, wie wirksam die normativ entwickelten didaktischen Überlegungen für Musikunterricht sind (Puffer & Hofmann, 2017), bedeutet das auch für die Modellierung von Erklärqualität, dass ein Rekurs auf fachübergreifende Bezüge notwendig wird.

17 Dahingehend fachspezifische Adaptionen werden auch für andere ästhetische Fächer als notwendig erachtet, was Befunde für den Kunstunterricht der Primarstufe (z. B. Schmidt, 2016) oder für den Sportunterricht (z. B. Herrmann & Gerlach, 2020) nahelegen.

Anhand der bisherigen Ausführungen zeigt sich, dass Modelle und Merkmale zur Beschreibung von Unterrichtsqualität sowohl aus fachübergreifender als auch fachspezifischer Perspektive entwickelt werden können – eine Forderung, die innerhalb der Unterrichtsforschung immer deutlicher gestellt wird (Praetorius, Herrmann et al., 2020; siehe auch Klieme, 2006; Terhart, 2013). Daran beteiligt sind vor allem zwei Protagonist*innen, deren Verhältnis zueinander noch nicht ausreichend geklärt ist: Erziehungswissenschaft und Fachdidaktik. Denn es variieren die Vorstellungen darüber, ob und inwieweit Verbesserungen des Lehrens und Lernens eher aus einer fachübergreifenden oder aus einer fachspezifischen Forschungsperspektive heraus beeinflusst werden können.

Während beispielsweise Schilcher und Krauss (2019) anhand von Beispielen aufzeigen, dass die Vermittlung von fachlichen Inhalten immer an diese gebunden sei und daher nicht aus einer fachübergreifenden Perspektive gedacht werden könne (siehe dazu auch Leinhardt, 2001), sind Kirchner et al. (2017) davon überzeugt, dass es allgemeine Prinzipien des Lehrens und Lernens gebe, die grundsätzlich auf die unterschiedlichen Inhalte aller Fächer angewendet werden können. Zudem werden verschiedene Forderungen an die Unterrichtsforschung gerichtet, die mit diesen unterschiedlichen Sichtweisen in Zusammenhang stehen. Dabei sollte der Fachunterricht einerseits stärker aus der jeweiligen fachdidaktischen Perspektive gedacht werden (Leuders, 2015; Shulman, 1986). Andererseits empfiehlt Terhart (2013) den einzelnen Fachdidaktiken, sich der Methoden und Theorien zu bedienen, die aus dem Feld der Bildungswissenschaft generiert wurden, aber das Ziel eigenständiger Fragestellungen zu verfolgen.

Daran knüpft die Forderung an, das Wissen über allgemeine Lehr-Lernprinzipien für einzelne Fachdidaktiken nutzbar zu machen (Cramer & Schreiber, 2018). Zudem wird der transdisziplinären Zusammenarbeit verschiedener Fachdidaktiken, wie beispielsweise bei den Forschungsprojekten DESI (siehe DESI-Konsortium, 2008), FALKO (Krauss et al., 2017) oder FALKE-q¹⁸ (**F**achspezifische **L**ehrer**k**ompetenzen im **E**rklären – **Q**ualität; Schilcher et al., 2021), bisher kaum abgerufenes Potenzial bescheinigt (Praetorius & Nehring, 2020). Dabei sei jedoch zu berücksichtigen, dass die Qualität von Unterricht aus Sicht der einzelnen Fächer „nicht nur von den gesetzten Zielen, sondern auch von den thematisierten Inhalten und den eingesetzten Methoden abhängt“, weshalb „alleine aus diesem Grund Unterschiede zwischen Fachdidaktiken in der Auseinandersetzung mit Unterrichtsqualität zu erwarten sind“ (Praetorius & Nehring, 2020, S. 298). Im Rahmen dieser Studie wird das gute Erklären einer Musiklehrkraft genau innerhalb dieses Spannungsfeldes verhandelt, da einerseits fachübergreifend relevante Merkmale und Charakteristika (siehe Kap. 2.3), andererseits jedoch auch fachspezifische Besonderheiten erwartet werden (siehe Kap. 2.4).

18 In der hier vorliegenden Quelle (Schilcher et al., 2021) ist von FALKE und nicht von FALKE-q die Rede. Die Spezifizierung folgt einer Kategorisierung des Forschungsprogramms FALKE in FALKE-q (Qualität), FALKE-e (Expertise) und FALKE-d (Digitalisierung).

Messverfahren

Ein zentrales Anliegen empirischer Unterrichtsforschung ist die Beantwortung der Frage, wie Unterrichtsqualität gemessen werden kann. Da es jedoch aufgrund der Komplexität von Unterricht nahezu unmöglich ist, dessen Qualität ganzheitlich zu erfassen, werden häufig nur einzelne Merkmale in den Blick genommen (z. B. Klieme & Rakoczy, 2008; Kunter et al., 2011). In diesem Zusammenhang konzentriert sich die folgende Darstellung vor allem auf die am Unterricht beteiligten Akteurinnen und Akteure: Lehrkräfte und Schüler*innen.

Ein Verfahren zur Messung von Unterrichtsqualität ist beispielsweise, Aussagen über guten Unterricht anhand des Zusammenhangs zwischen dem professionellen Wissen von Lehrkräften (siehe Kap. 2.2.2) und Veränderungen in der Leistung von Schüler*innen abzuleiten (z. B. Kunter et al., 2011). Ausgangspunkt dieses Vorgehens ist die Orientierung am Output von Unterricht, „also an tatsächlich erreichten Effekten und Wirkungen“ (Terhart, 2002, S. 104) wie dem nachgewiesenen Wissenszuwachs von Schüler*innen. Sowohl die Daten der Meta-Metaanalyse von Hattie (z. B. Hattie et al., 2014; siehe dazu auch Lotz & Lipowsky, 2015) als auch der COACTIV-Studie zum Professionswissen von Mathematiklehrkräften (Kunter et al., 2011) zeigen, dass der Grad an kognitiver Aktivierung und konstruktiver Unterstützung im Unterricht nicht nur mit dem fachdidaktischen Wissen der Lehrkraft, sondern auch mit dem Lernerfolg der Schüler*innen zusammenhängt. Beide Merkmale guten Unterrichts fungieren dabei als Mediator und lassen den Schluss zu, dass das fachdidaktische Wissen einer Lehrkraft für den Lernerfolg von Schüler*innen prädiktiv valide ist (Baumert & Kunter, 2011).¹⁹ Diese positiven Aspekte des Unterrichts wiederum führen zu einem höheren Lernerfolg bei den Schüler*innen und nehmen damit die Rolle von Mediatoren ein, da sie das Potenzial auf Seiten der Lehrkräfte mit der Lernleistung auf Seiten der Schüler*innen positiv in Beziehung setzen (Baumert & Kunter, 2011). Im Zusammenhang dieser Ergebnisse bleibt jedoch unklar, welcher Anteil dieser prädiktiven Validität auf das Erklärungswissen zurückzuführen ist.

Allerdings gibt es auch viele Testverfahren, die Veränderungen im Wissen beziehungsweise in der Leistung entweder ausschließlich bei Schüler*innen oder Lehrkräften diagnostizieren. Mit Hilfe von länderübergreifenden Schulleistungsstudien²⁰ wie DESI (z. B. DESI-Konsortium, 2008), IGLU (z. B. Hußmann et al., 2017), TIMSS (z. B. Schwippert et al., 2020) oder PISA (z. B. Reiss et al., 2019) werden in regelmäßigen Abständen die Leistungen von Schüler*innen gemessen und in den dazugehörigen Begleitforschungen untersucht. Die genannten Studien beschränken sich jedoch auf die mathematisch-naturwissenschaftlichen Fächer (PISA, TIMSS, IGLU-E) sowie auf die Fächer Deutsch, mit Schwerpunkt auf den Bereich Lesen (DESI, IGLU, IGLU-E, PISA), und

19 Prädiktive Validität bedeutet in diesem Fall: Je höher das Fachdidaktische Wissen einer Lehrkraft ist, desto mehr Potenzial an beispielsweise kognitiver Aktivierung, individueller Lernunterstützung sowie der Passung zwischen den verwendeten Aufgaben und den Anforderungen des Lehrplans kann in ihrem Unterricht festgestellt werden.

20 Das von Weinert (1999) entwickelte Rahmenmodell zur Kompetenzorientierung hatte eigentlich zum Ziel, den Begriff „Leistung“ durch „Kompetenz“ zu ersetzen. Trotzdem hat sich durchgesetzt, von Schulleistungsstudien zu sprechen.

Englisch (DESI). Für das Fach Musik fehlen solche Large-Scale-Erhebungen bisher ebenso wie ein breit akzeptiertes Kompetenzmodell beziehungsweise „ausreichende Erkenntnisse über notwendige Kompetenzausprägungen“ (Hasselhorn & Knigge, 2018, S. 199).

Forschung zur musikbezogenen Kompetenz von Schüler*innen gibt es innerhalb der Musikpädagogik erst in einzelnen Teilbereichen. Mit dem KOPRA-M-Test entwickelte und validierte Hasselhorn (2015) sowohl ein Testinstrument als auch das dazugehörige Kompetenzmodell zu musikpraktischen Kompetenzen in den „drei Dimensionen Singen, instrumentales Musizieren und Rhythmusproduktion“ (Hasselhorn & Lehmann, 2014, S. 89) von Schüler*innen. Auch zur Erfassung der Kompetenz *Musik wahrnehmen und kontextualisieren* wurden sowohl Testaufgaben (Knigge, 2011) als auch ein dazugehöriges Kompetenzmodell entwickelt und validiert (Jordan et al., 2012; Niessen et al., 2008). Der KoMus-Test misst neben der Wahrnehmungsfähigkeit in Form musikalischer Gedächtnisleistung von Schüler*innen der 6. Jahrgangsstufe auch die in drei Dimensionen unterteilte Kompetenz, musikalisches Wissen auf Basis des Wahrgenommenen anzuwenden: Die Fähigkeit zur Verbalisierung, zur Notation und zur historisch-kulturellen Kontextualisierung von Musik (Jordan et al., 2012; Niessen et al., 2008). In beiden Modellen werden die jeweiligen Kompetenzen der Schüler*innen anhand geltender Curricula abgeleitet, wobei ein solch „pragmatisches Vorgehen“ durch einen „allgemeingültigen Katalog musikalischer Bildungsziele“ (Hasselhorn & Knigge, 2018, S. 199) obsolet würde.

Es ist mit diesen Instrumenten möglich, Aspekte musikalischer Kompetenzen von Schüler*innen zu messen. Unklar ist in diesem Zusammenhang jedoch, welche Rolle schulischer Musikunterricht bei der Entwicklung musikalischer Kompetenzen einnimmt und welchen Einfluss darauf Musiklehrkräfte haben. Demzufolge liegen auch keine Erkenntnisse darüber vor, ob das Erklären einer Musiklehrkraft dazu beiträgt, die genannten Kompetenzen zu fördern. Allerdings wird die Orientierung am Test-Output von Schüler*innen hinsichtlich ihrer musikbezogenen Kompetenzen als Antwort auf die Frage nach der Qualität von Musikunterricht auch kritisiert. Skepsis herrscht vor allem in Bezug auf das Messen ästhetischer Aspekte von Musik sowie hinsichtlich der damit verbundenen Standardisierung oder Quantifizierung (Rolle, 2008). Zudem wird in Frage gestellt, inwieweit musikbezogene Bildung in Form von Kompetenzen überhaupt beschrieben werden kann (Vogt, 2008).

Um die professionelle Kompetenz von Lehrkräften zu messen (siehe ausführlich Kap. 2.2.2), wurden in vielen Unterrichtsfächern Professionswissenstests entwickelt und validiert. Als Vorreiter im deutschsprachigen Raum²¹ können die Kompetenztests für Mathematiklehrkräfte aus den Forschungsprojekten TEDS-M (Blömeke et al., 2010b) und COACTIV (Kunter et al., 2011) gelten, eine ausführliche Darstellung von Professionswissenstests verschiedener Fächer ist an anderer Stelle zu finden (z. B. Krauss et al., 2017). Für die Musikpädagogik entwickelten und validierten Puffer und Hofmann (2017) mit FALKO-M den ersten Professionswissenstest für (angehende) Musiklehrkräfte, der Fachwissen und fachdidaktisches Wissen valide und getrennt voneinander messen kann.

21 Aus dem amerikanischen Raum ist beispielsweise Ball et al. (2005) zu nennen.

Während in anderen Fachkontexten wie Mathematik bereits gezeigt werden konnte, dass eine Lehrperson, je mehr sie über professionelles Wissen verfügt, die Kompetenzentwicklung von Schüler*innen positiv beeinflusst (Kunter et al., 2011), liegen für den Musikunterricht in Deutschland diesbezüglich noch keine Erkenntnisse vor. Neben Werkzeugen zur Messung der musikalischen Kompetenz von Schüler*innen, wie dem KOPRA-M-Test oder dem KoMus-Test, bedarf es „zur Untersuchung von Wirkzusammenhängen rund um schulischen Musikunterricht“ (Puffer & Hofmann, 2017, S. 279) auch Instrumente wie den FALKO-M-Test, die das Professionswissen oder -handeln von Musiklehrkräften messen können. Der Befund, dass es innerhalb eines musikpädagogischen Forschungsvorhabens noch nicht gelungen ist, die unterschiedlichen Wirkzusammenhänge aus Lehrperson, Unterrichtsangebot und Nutzung durch Schüler*innen zu kombinieren (Puffer & Hofmann, 2017), hat nach wie vor Gültigkeit.

Während auf Basis des Outcomes von Unterricht Aussagen über dessen Wirkung und damit auch dessen Qualität getroffen werden können (Terhart, 2002), wird das Vorgehen, ausschließlich das Wissen der Lehrkräfte zu messen und auf Basis dessen auf Unterrichtsqualität zu schließen, hinsichtlich der Komplexität von Unterricht als verkürzt kritisiert (Blömeke, König et al., 2015). Blömeke, Gustafsson et al. (2015) merken dazu an, dass das über schriftliche Tests gemessene Wissen von Lehrkräften lediglich Auskunft über ihre Leistungspotenziale und nicht über ihre tatsächliche Realisation von Unterricht geben kann. Es stellt nach wie vor eine große Herausforderung dar, die Qualität von Unterricht anhand der Testergebnisse von Kompetenztests abzuleiten. Dabei ist nicht nur unter großem Aufwand zu gewährleisten, einen Zusammenhang zwischen der Leistung von Schüler*innen und dem Wissen einer Lehrkraft herstellen zu können. Als problematisch wird auch erachtet, anhand der Testergebnisse über das Wissen von Lehrkräften auf Unterrichtsqualität zu schließen, die auch Handlungskompetenzen der Lehrpersonen berücksichtigen sollte (Blömeke, König et al., 2015).

Beurteilungsverfahren

Neben Leistungsmessung und Wissenstests werden als weitere Möglichkeit auch Beurteilungen von Unterricht durch Lehrkräfte, Schüler*innen oder weitere relevante Personengruppen wie Fachdidaktiker*innen genutzt, um auf Basis von deren Aussagen und Einschätzungen auf Unterrichtsqualität zu schließen. Zur Anwendung kommt dieses Vorgehen unter anderem in den Schulleistungstests PISA, IGLU und DESI. Beispielsweise wurden in der DESI-Studie bei über 10.000 Schüler*innen der neunten Jahrgangsstufe nicht nur deren (sprachliche) Leistung in den Fächern Deutsch und Englisch gemessen, sondern auch deren Einschätzungen bezüglich der Qualität von Unterricht abgefragt – ebenso wie von Eltern, Schulleitungen, Lehrkräften oder Hochschullehrenden (DESI-Konsortium, 2008).

In Bezug auf die Einschätzungen externer Beobachterinnen und Beobachter, die beispielsweise an der Ausbildung von angehenden Lehrkräften beteiligt sind, stellt Praetorius (2013) fest, dass diese innerhalb der Unterrichtsforschung zwar „oft als besonders geeignet zur Messung von Unterrichtsqualität beschrieben“ werden, sich dabei „jedoch

diverse Einschränkungen von Beobachterratings“ (S. 181) zeigen. Während diesbezüglich keine Gefahr für den Einfluss sozialer Erwünschtheit auf die Beurteilungen besteht – dies kann sich beispielsweise bei Lehrkräften zeigen, die ihren eigenen Unterricht beurteilen sollen (Kleickmann et al., 2019) – nennt die Autorin die eingeschränkte Informationsgrundlage bezüglich der zu beurteilenden Situation als Nachteil (Praetorius, 2013). Lehrkräften wiederum würden diese Kontextinformationen vorliegen. Beiden Gruppen wird jedoch eine hohe pädagogische wie fachdidaktische Expertise zugeschrieben, die beispielsweise der Gruppe von Schüler*innen fehlt (Kleickmann et al., 2019).

Daher wird bei der Beurteilung von Unterrichtsqualität durch Schüler*innen mitunter die Güte ihrer Urteile in Frage gestellt (Clausen, 2002; Göllner et al., 2016; Kleickmann et al., 2019). Neben der fehlenden pädagogischen und didaktischen Expertise von Schüler*innen (Göllner et al., 2016) lassen sich auch Einflüsse auf ihre Wahrnehmung in Form des Halo-Effekts²² (Clausen, 2002) oder des Dr.-Fox-Effekts²³ (Kleickmann et al., 2019) feststellen. Ein weiteres kritisches Argument ist die subjektive Involviertheit der Lernenden ins Unterrichtsgeschehen (Göllner et al., 2016) – ein Aspekt, der jedoch auch auf Lehrkräfte zutrifft, die Aussagen zu ihrem eigenen Unterricht treffen sollen. Entgegen dieser Kritik zeigen jedoch die Daten der DESI-Studie, dass Urteile von Schüler*innen hinsichtlich zweier unterschiedlicher Fächer – in diesem Fall Deutsch und Englisch – auch weitgehend unabhängig voneinander getroffen werden (Wagner et al., 2013), was für deren Güte spricht. Weiterhin zielen alle unterrichtlichen Bemühungen einer Lehrkraft auf Schüler*innen als Adressat*innen, weshalb dadurch eine Perspektive abgedeckt wird, die nur Schüler*innen einnehmen können (Kleickmann et al., 2019).

Aus diesen Gründen kann die Validität der Beurteilungen von Schüler*innen durchaus angenommen werden (Wagner et al., 2013). In diesem Zusammenhang konnten Wisniewski, Zierer et al. (2020) hinsichtlich ihres Fragebogens zum teaCh-Modell zeigen, dass Schüler*innen der Sekundarstufe I und II sowohl niedrig- wie hoch-inferente Items beantworten können, die jedoch alle den Anspruch haben, Tiefenmerkmale von Unterricht zu beschreiben. Generell ist in diesem Zusammenhang zu bedenken, dass die sprachliche Anforderung bei der Formulierung der Items für Schüler*innen zu berücksichtigen (Göllner et al., 2016) und eine ausführliche Pilotierung angezeigt ist (Bühner, 2021, S. 91; für die Anwendung im Projekt FALKE-q-Mu siehe Frei, 2019). Lenske (2016) konnte beobachten, dass Schüler*innen der Primarstufe manche Begriffe in den Fragebogenitems abweichend vom eigentlich intendierten Sinn verstanden haben.

Weiterhin gibt es verschiedene Arbeiten, die Einschätzungen von Schüler*innen aus Primar- und Sekundarstufe mit denen von Lehrkräften hinsichtlich der Qualitätsmerkmale von Unterricht vergleichen (Clausen, 2002; Fauth et al., 2014; Kunter & Baumert, 2006). Im Sekundarbereich konnte gezeigt werden, dass zwischen beiden Gruppen nur

22 Der *Halo-Effekt* bezeichnet eine kognitive Verzerrung der Wahrnehmung, indem von einzelnen (bekannten) Eigenschaften einer Person auf weitere (unbekannte) geschlossen und diese somit „überlagert“ werden (Clausen, 2002, S. 188).

23 Der *Dr.-Fox-Effekt* beschreibt, dass eine Gruppe von Adressat*innen über inhaltliche Mängel eines Vortrags aufgrund selbstbewusster Rhetorik der Vortragenden Person hinweggetäuscht wurden (Helmke, 2017). Im übertragenen Sinne könnten Schüler*innen die Qualität einer Erklärung durch ihre Lehrkraft überschätzen und ihre Beurteilung verzerrt werden.

wenig Übereinstimmung in der Beurteilung von Unterrichtsqualität festzustellen ist und für die Bewertung beider Gruppen unterschiedliche Merkmale relevant sind (Clausen, 2002; Kunter & Baumert, 2006). Daraus kann jedoch nicht geschlossen werden, dass eine Perspektive einer anderen vorzuziehen sei, sondern vielmehr beide Sichtweisen gleichermaßen Berechtigung als Informationsquelle haben und für ein umfassenderes Bild von Unterrichtsqualität notwendig sind (Clausen, 2002; Kunter & Baumert, 2006). Während beispielsweise die Beurteilung des Anteils kognitiver Aktivierung im Unterricht von Schüler*innen der Sekundarstufe eine größere Nähe zur tatsächlichen Unterrichtswirklichkeit aufweisen kann und damit genauer zu sein scheint, eignen sich die Angaben der Lehrkräfte eher dazu, die Auswahl und Implementierung von Aufgaben besser zu bewerten (Kunter & Baumert, 2006).

Fauth et al. (2014) konnten für den Primarbereich ähnliche Befunde feststellen, wobei ein entscheidender Unterschied zu den beiden Studien aus dem Sekundarbereich auf methodischer Ebene auszumachen ist: Gefragt wurde dort nicht nach einer generellen Beurteilung von Unterricht, sondern nach der Einschätzung zu einer konkreten Unterrichtseinheit beziehungsweise -situation. Dies wurde von den Autor*innen in der Rückschau auf die Studie positiv beurteilt, da die Grundschüler*innen ihre Einschätzungen auf Grundlage eines klar eingegrenzten Gegenstandes abgeben konnten (Fauth et al., 2014). Allen genannten Studien zur Beurteilung durch Schüler*innen und Lehrkräften ist gemein, dass beide Gruppen Unterricht beurteilen, an dem sie selbst beteiligt waren (Clausen, 2002; Fauth et al., 2014; Kunter & Baumert, 2006, Wagner et al., 2013). Bislang sind keine Studien mit Beurteilungsverfahren bekannt, in denen Unterricht von Personen beurteilt wird, die selbst nicht daran teilgenommen haben. Es besteht die Annahme, dass dadurch der Kritik subjektiver Involviertheit begegnet werden könnte. Unklar ist bislang jedoch, ob eine solche Beurteilung fremden Unterrichtsgeschehens tatsächlich zu objektiveren und damit valideren Ergebnissen kommen kann.

Obwohl die Einschätzungen über die Qualität von Unterricht zwischen Schüler*innen sowie Lehrkräften häufig nicht übereinstimmen und noch Unklarheit bezüglich des Umgangs mit dieser Tatsache herrscht (Praetorius et al., 2018), wird trotzdem empfohlen, verschiedene Perspektiven zu erheben, um dadurch ein umfassenderes Bild von Unterrichtsqualität zu erhalten (Clausen, 2002; Göllner et al., 2016; Kunter & Baumert, 2006). Dieses Vorgehen wurde beispielsweise in einer Untersuchung zum guten Geschichtsunterricht angewendet, indem Einschätzungen von Schüler*innen, Lehrkräften und Hochschullehrenden erhoben und miteinander verglichen wurden, wobei alle Perspektiven mit der gleichen Gewichtung für die Bestimmung der Unterrichtsqualität in Betracht gezogen wurden (Gautschi, 2009). Bei der Darstellung des methodischen Vorgehens bezieht sich die Arbeit auf Flick (2009), der das Einbeziehen verschiedener Perspektiven auf unterschiedliche Aspekte von Forschung mit dem Begriff *Triangulation* (z. B. Datentriangulation oder Methodentriangulation) umschreibt. Obwohl der Autor bei dieser Zusammenschau nicht explizit erwähnt, dass auch subjektive Einschätzungen verschiedener am Unterricht Beteiligter zum selben Gegenstand (nämlich Unterricht) miteinander trianguliert werden können, ähnelt die *Triangulation von Perspektiven*, wie es Gautschi (2009)

beschreibt, den anderen Verfahren von Methoden- und Datentriangulation (siehe Flick, 2009).

Auch innerhalb der Musikpädagogik kommen vermehrt Beurteilungsverfahren zum Einsatz, um die Vorstellungen über guten Musikunterricht von verschiedenen Gruppen zu erfassen. Eine erste musikpädagogische Studie, in der Schüler*innen systematisch befragt wurden hat sich mit deren Vorlieben bezüglich der im Musikunterricht verwendeten Methoden auseinandergesetzt (Schulten, 1980). Mit der MASS-Studie führte Heß (2011) eine umfassende Untersuchung zur Sichtweise von Schüler*innen der Sekundarstufe²⁴ auf Musikunterricht durch und konnte auf Basis einer Fragebogenerhebung unter anderem zeigen, dass Musikunterricht „keineswegs zu den unbeliebten Fächern“ (Heß, 2011, S. 23) gehöre, jedoch vor allem bei den Schüler*innen beliebt sei, die ein Instrument spielen. Auch wurde auf Basis von Beurteilungen durch Schüler*innen ein Modell zur Qualität von Musikunterricht entworfen (Hafen et al., 2011). Das Projekt ProBiNi (Haas et al., 2019) untersuchte anhand eines Fragebogens, inwieweit sich die Urteile von Schüler*innen zu den in der bildungswissenschaftlichen Forschung prominenten und wirkmächtigen drei Grunddimensionen guten Unterrichts zwischen Musikklassen- und normalem Musikunterricht unterscheiden. Die Autor*innen konnten dabei unter anderem zeigen, dass sich ein Modell von Unterrichtsqualität, das auf diesen drei Basisdimensionen fußt, „sowohl in den Musikklassen als auch in den Vergleichsklassen äquivalent abbilden lässt“, jedoch die Schüler*innen der Musikklassen im Vergleich „einen signifikant höheren Grad an Klassenführung im Musikunterricht wahrnehmen“ (Haas et al., 2019, S. 149–150). Neben solchen Studien, die anhand von Fragebögen und meist geschlossenen Itemformaten Beurteilungen von Schüler*innen erheben, rekonstruierten andere Untersuchungen anhand von (Gruppen-)Interviewformaten auf Basis der Äußerungen von Schüler*innen bedeutsame Prinzipien von Musikunterricht (z. B. Göllner, 2017). Harnischmacher und Hofbauer (2013) kombinierten offene und geschlossene Antwortformate innerhalb eines Fragebogens zur Beurteilung von Musikunterricht aus der Perspektive von Schüler*innen aus Sekundarstufe I und II.

Beurteilungen durch Musiklehrkräfte nehmen ebenfalls eine wichtige Rolle ein, auf deren Basis (Qualitäts-)Aspekte von Musikunterricht beschrieben werden. Allerdings kommen auch hier unterschiedliche Erhebungsinstrumente und -formen zum Einsatz. Harnischmacher et al. (2017) konnten anhand einer Fragebogenstudie mit Musiklehrkräften zeigen, dass der Grad an im Unterricht vorliegender Kompetenzorientierung – genauer inwieweit Musiklehrende ihren Schüler*innen den Aufbau von Kompetenzen ermöglichen – durch Motivation, Selbstreflexion und Musizierpraxis vorhergesagt werden kann, wobei der letzte Aspekt als Ausdruck von Professionswissen verstanden und von den Autor*innen diesbezüglich weiterer Forschungsbedarf reklamiert wurde. Musikpädagogische Interviewstudien mit (angehenden) Lehrkräften folgen dabei unter anderem explorativen Fragestellungen. Godau (2018) befragte Referendar*innen sowie junge Lehrkräfte in Einzel- sowie Gruppeninterviews nach ihrem Entscheidungsverhalten bezüglich des Einsatzes von *Dingen* im Musikunterricht, wie beispielsweise Instrumente

24 Für den Primarbereich legte Gaul (2009) eine Arbeit vor und setzte sich darin mit Beurteilungen über Musikunterricht aus Schülersicht auseinander.

oder Musikapps. Um zu rekonstruieren, wie Musiklehrkräfte über ihr eigenes Unterrichten von Musiktheorie nachdenken, untersuchte Ruf (2014) bayerische Musiklehrkräfte. Auf Basis einer Interviewstudie konnte er zeigen, dass nicht nur unterschiedliche Begriffe von Musiktheorie, sondern diesbezüglich auch verschiedene Ziele der Lehrenden vorliegen.

Das Prinzip der Triangulation mehrerer Perspektiven liegt neben Studien, die bisher schon erwähnt wurden – Göllner (2017) verglich die Perspektiven von Schüler*innen und Lehrkräften in Bezug auf Bläserklassenunterricht anhand von Interviewaussagen miteinander –, auch weiteren musikpädagogischen Forschungsarbeiten zugrunde. Niessen und Lehmann-Wermser (2006) befragten Musiklehrkräfte vor und nach einer Musikstunde bezüglich ihrer Planung und Durchführung. Sie verglichen die gewonnenen Aussagen mit den Interviewaussagen, welche die Schüler*innen über ihre Motivation und ihren Lernerfolg in der jeweiligen Unterrichtsstunde trafen, und konnten unter anderem ein Missverhältnis beider Perspektiven in Bezug auf die Relevanz von Lerninhalten feststellen (Niessen & Lehmann-Wermser, 2006). Unterschiedliche Wahrnehmungsdimensionen auf Musikunterricht wurden aus Sicht von Schüler*innen, Studierenden und (angehenden) Lehrkräften anhand eines videobasierten Tests untersucht (Harnischmacher & Hofbauer, 2011). Anhand dieser beiden am Unterricht beteiligten Gruppen untersuchte Weidner (2018a) die jeweilige Sichtweise auf Unterrichtsgespräche und Niessen (2010) die beiden Perspektiven hinsichtlich der Bedeutung von Verarbeitungstiefe im Musikunterricht.

Zusammenfassend zeigt sich für den (musikpädagogischen) Diskurs um Unterrichtsqualität, dass sowohl Mess- als auch Beurteilungsverfahren zur Anwendung kommen. Dabei werden teilweise verschiedene Perspektiven eingeholt, um auf Basis der unterschiedlichen Aussagen ein umfassenderes Bild über Musikunterricht zu erhalten und anhand dessen auch Qualitätsmerkmale ableiten zu können. Während die Perspektive von Musiklehrkräften vor allem in qualitativen Verfahren eingeholt wird – eine Ausnahme bilden beispielsweise Harnischmacher und Hofbauer (2011) –, kamen bisher keine Erhebungsinstrumente zum Einsatz, die neben Schüler*innen, Studierenden und Lehrkräften auch weitere externe Expert*innen wie Hochschullehrende in die Erhebung mit einbeziehen. Dies gilt sowohl hinsichtlich der Qualität von Musikunterricht als auch bezüglich der Rolle des Erklärens im Fach Musik.

Paradigmen der Forschung über Lehrkräfte

Die bisherigen Ausführungen nahmen verschiedene Komponenten und Merkmale guten Unterrichts sowie Möglichkeiten des Messens und Beurteilens von Unterrichtsqualität in den Blick. An einigen Stellen wurde bereits deutlich, dass die Lehrkraft dabei eine zentrale Rolle einnimmt, weshalb sich nun der Blick auf die Forschung über Lehrkräfte richtet. Diese kann auf eine lange Tradition zurückblicken und folgt unterschiedlichen Paradigmen²⁵, die sich forschungsmethodisch zum Teil erheblich unterscheiden (Tillmann,

25 Zum Begriff *Paradigma* siehe z. B. Kiel, 2014, S. 937.

2014). In ihrer Übersicht zu Paradigmen, die der pädagogisch-psychologischen Forschung über Lehrkräfte zuzuordnen sind, verweisen Krauss und Bruckmaier (2014) allerdings auch auf die „theoretischen Überschneidungen zwischen den Paradigmen“, weshalb die „Paradigmenwechsel nicht als echte Brüche zu sehen, sondern eher als Akzentverschiebungen und Weiterentwicklungen“ (S. 242) zu deuten sind. In der Weiterentwicklung des Modells ergänzt Krauss (2020) dieses um den Ansatz professioneller Kompetenz, der als aktuell zentrales Paradigma eingeordnet wird und dem sich die vorliegende Arbeit verpflichtet. Für dessen ausführliche Darstellung ist es jedoch zunächst notwendig, verschiedene Ansätze der Forschung über Lehrkräfte sowie deren jeweilige (Weiter-)Entwicklungen zu entfalten, um dabei auch einen gesonderten Blick auf den musikpädagogischen Diskurs richten zu können.²⁶

In der ersten Hälfte des 20. Jahrhunderts wurde versucht, die Lernerfolge von Schüler*innen mit den Persönlichkeitseigenschaften von Lehrkräften zu erklären, weshalb dieser Ansatz mit dem Begriff *Persönlichkeits-Paradigma* bezeichnet wird (Krauss & Bruckmaier, 2014). Während sich dieses Paradigma aus erziehungswissenschaftlicher Sicht jedoch als unzureichend erwies, weil im Zuge dessen weder das Unterrichtshandeln einer Lehrkraft noch deren direkte Interaktion mit Schüler*innen berücksichtigt wurde (Helmke, 2017), prägt er den musikpädagogischen Diskurs weiterhin in relevanter Weise (Hofmann, 2011; siehe z. B. Oberhaus, 2007).²⁷

Als Reaktion auf diese zwei Kritikpunkte des *Persönlichkeits-Paradigmas* etablierten sich in der erziehungswissenschaftlichen Forschung zunächst zwei Ansätze (Krauss & Bruckmaier, 2014). Während sich der Fokus anfangs vor allem auf das Verhalten der Lehrkraft und im Zusammenhang damit auf den Unterrichtsprozess richtet (*Prozess-Produkt-Paradigma*), werden später auch Merkmale von Schüler*innen, wie deren individuelle Informationsverarbeitungsprozesse, als Mediator (*Prozess-Mediations-Produkt-Paradigma*) zwischen Unterrichtsprozess und -ertrag berücksichtigt (Krauss & Bruckmaier, 2014). Dabei wurden als *Prozesse* von Unterricht beispielsweise die Anzahl an Lehrerfragen und als dessen *Produkte* zum Beispiel der Lernzuwachs der Schüler*innen empirisch erfasst, um in einem weiteren Schritt „die Berechnung von Zusammenhangsmaßen (etwa Korrelationen) zwischen den untersuchten Prozessen und Produkten“ (Krauss & Bruckmaier, 2014, S. 241) zu ermöglichen. In der Folge wurde die Frage nach der guten Lehrperson durch „Erkenntnisse aus der kognitionspsychologischen Expertiseforschung“ (Krauss, 2020, S. 154) beeinflusst, deren Konzept auf der Orientierung an besonders erfolgreichen Personen einer Domäne beruht (Mulder & Gruber, 2011). Dadurch richtet sich der Fokus wieder auf die Person der Lehrkraft, ohne jedoch Charaktereigenschaften wie beim *Persönlichkeits-Paradigma* in den Vordergrund zu stellen, sondern sich im Gegensatz dazu für das Wissen und Können von Lehrkräften zu interessieren (Krauss, 2020).

Prägend für die Sichtweise, die Befähigung zum Unterrichten als Expertise anzusehen, waren im deutschsprachigen Raum zunächst vor allem die Ausführungen Brommes

26 Auch im Rahmen der Perspektive von Lehrer*innenbildung werden verschiedene Ansätze unterschieden, siehe dazu Cramer et al., 2020.

27 Diskutiert wird beispielsweise das Spannungsfeld zwischen Künstler*in und Pädagog*in, das im Bereich der Instrumentalpädagogik von Bedeutung ist (z. B. Grimmer & Lessing, 2008; Rübke, 2018).

(1992). Da für dieses Paradigma unter anderem die Unterscheidung zwischen Expert*innen und Noviz*innen ein Charakteristikum darstellt (Gruber & Stöger, 2011) und es als wichtige Aufgabe der Expertiseforschung gilt, diese Unterschiede herauszuarbeiten (Mulder & Gruber, 2011), trägt es zur konzeptuellen Rahmung verschiedener Studien hinsichtlich der Messung von Lehrkräftekompetenzen bei. Untersucht werden in diesem Zusammenhang beispielsweise die Fragen, inwieweit sich angehende und erfahrene Lehrkräfte in ihrem professionellen, für die Ausübung des Berufs notwendigen, Wissen unterscheiden, oder ob es Wissenszuwächse bei steigender beruflicher Erfahrung gibt (z. B. Baumert & Kunter, 2006; Blömeke et al., 2010a; Krauss et al., 2017; Kunter et al., 2011).

In der musikpädagogischen Forschung spielt der Expert*innen-Noviz*innen-Ansatz unter anderem in verschiedenen Studien zur Messung von Wissen, Können und (Reflexions-)Fähigkeiten von Lehrkräften eine Rolle (z. B. Harnischmacher & Hofbauer, 2011; Hofbauer, 2017; Puffer & Hofmann, 2017). Die Ergebnisse aus FALKO-M zeigen beispielsweise, dass sich Musik-Lehramtsstudierende und (angehende) Musiklehrkräfte hinsichtlich ihres Wissens um musikbezogene Instruktionsstrategien im Schnitt mit einer mittleren bis hohen Effektstärke (Cohens $d = 0,71$) unterscheiden (Puffer & Hofmann, 2017). Mit Hilfe von Professionstests wie FALKO-M verfügt die Musikpädagogik über erste Werkzeuge, um Unterschiede zwischen Expert*innen und Noviz*innen herauszuarbeiten und damit gleichzeitig ein Hauptanliegen der Expertiseforschung zu erfüllen.

Einen ebenfalls empirischen, allerdings nicht zwangsweise pädagogisch-psychologischen Zugang verfolgt der berufsbiografische Ansatz. Nach Terhart (2011) fokussiert dieser eine individualisierte Sichtweise auf Lehrkräfte und deren (Weiter-)Entwicklung über ihre gesamte berufliche Laufbahn hinweg. Inhaltliche Themen sind diesbezüglich beispielsweise die Berufswahlentscheidung von Lehrkräften (z. B. Rothland, 2014) oder deren Einstiegsphase in den Beruf (z. B. Keller-Schneider & Hericks, 2014), wobei das zentrale Anliegen ebenfalls das Entwickeln von Expertise ist (Terhart, 2011).

Innerhalb der Musikpädagogik ist die Bedeutung dieses Forschungsparadigmas durch verschiedene Forschungsvorhaben dokumentiert, die sich unterschiedlichen Fragestellungen zuordnen lassen (Lessing & Stöger, 2018). Neben Beiträgen, die sich dem Berufsalltag von Musiklehrkräften widmen (z. B. Ruf, 2014), oder Studien, die verschiedene Phasen des Berufslebens von Musiklehrkräften (z. B. Bailer, 2009) oder Entscheidungen zur Berufswahl in den Blick nehmen (z. B. Neuhaus, 2008), spielt auch die Zufriedenheit im Beruf (z. B. Pfeiffer, 1994) sowie das bereits in der Ausbildung immanente Spannungsfeld zwischen der Persönlichkeit als Künstler*in sowie als Pädagog*in eine Rolle (z. B. Dreßler, 2017). Weiterhin können auch empirische Befunde über die Selbst- und Individualkonzepte von (fachfremd unterrichtenden) Musiklehrkräften (z. B. Hammel, 2011; Niessen, 2006b) diesem Forschungsparadigma zugeordnet werden. Untersuchungen zum unterrichtlichen Erklären von Musiklehrkräften unter Heranziehung ihrer berufsbiografischen Perspektive stellen bisher jedoch ein Desiderat dar.

Im Vergleich zu allen bisher genannten empirischen Paradigmen verfolgt der strukturtheoretische Bestimmungsansatz ein anderes forschungsmethodisches Vorgehen. Aus strukturtheoretischer Sicht (z. B. Helsper, 2014; Oevermann, 2002) wird der Lehrberuf als das Bewältigen komplexer und unsicherer Situationen des Unterrichts betrachtet

(Terhart, 2011). Entscheidend für die Weiterentwicklung einer Lehrkraft ist ihre Fähigkeit zur Reflexion des Unterrichts, wobei die Herausforderung vor allem darin besteht, innerhalb des Spannungsfeldes zwischen einer heterogenen Schüler*innenschaft und einheitlichen curricularen Vorgaben zu agieren (Terhart, 2011).

Auch dieser Forschungsansatz findet innerhalb des musikpädagogischen Schrifttums Berücksichtigung und blickt hinsichtlich einer umfassenden Bestandsaufnahme zur Struktur und den Rahmenbedingungen von Musikunterricht sowie der damit verbundenen Rolle von Musiklehrkräften bereits auf eine längere Tradition zurück (z. B. Jank et al., 1986; Jank, 2018). In jüngster Zeit werden gerade die Reflexionsfähigkeit und ihre Rolle in der Aus- und Weiterbildung fokussiert, wobei in diesem Zusammenhang auch empirische Methoden zum Einsatz kommen (z. B. Neuhaus, 2019; Prantl & Wallbaum, 2017). Eine Zuspitzung auf das Reflektieren von Erklärhandlungen im Kontext von Musikunterricht fand bisher innerhalb der musikpädagogischen Forschung noch keine Berücksichtigung, weshalb diesbezüglich weitere Vorhaben angezeigt wären.

Die bisherige Darstellung fußt auf mehreren Beiträgen aus der Forschungsliteratur, die teils unterschiedliche Kategorisierungen der Ansätze zur Professionalität von Lehrkräften vornehmen und dabei sowohl Gemeinsamkeiten als auch Unterschiede aufweisen (z. B. Krauss, 2020; Mulder & Gruber, 2011; Terhart, 2011). Während Terhart (2011) den berufsbiografischen neben den strukturtheoretischen Ansatz stellt, subsumieren Mulder und Gruber (2011) beide Paradigmen unter der Bezeichnung *Professionsforschung*. Unter dem Begriff *Professionalität* wiederum fasst Terhart (2011) auch den kompetenztheoretischen Ansatz, der bei Mulder und Gruber (2011) neben dem Professionsansatz und dem Expertiseansatz verortet wird. Krauss und Bruckmaier (2014) fokussieren pädagogisch-psychologische Paradigmen, wobei an dieser Stelle der kompetenztheoretische Ansatz noch ausgespart wird. In einer aktuellen Publikation ergänzt Krauss (2020) die Darstellung um das Paradigma der Professionellen Kompetenz und wertet dieses als „eine aktuelle Weiterentwicklung des Expertise-Paradigmas“ (S. 155).

Anhand dieser knappen Zusammenfassung zeigt sich, dass die bisherige Diskussion um Paradigmen der Forschung über Lehrkräfte noch zu keiner einheitlichen Kategorisierung und keinem gemeinsamen Verständnis von Begriffen geführt hat, wenngleich sich Konsenslinien erkennen und stetige Weiterentwicklungen beobachten lassen. So besteht größtenteils Einigkeit darin, dass das kompetenztheoretische Paradigma einen aktuell zentralen Ansatz darstellt, der nach Voraussetzungen für die Berufsausübung fragt (Mulder & Gruber, 2011; Krauss, 2020) und in diesem Zusammenhang „Kompetenzbereiche sowie Wissensdimensionen definiert“ (Terhart, 2011, S. 207). Auch wird die enge Verbindung zwischen Kompetenz- und Expertiseansatz an mehreren Stellen deutlich, ohne dabei auf Trennungslinien zu verzichten (Krauss, 2020; Krauss & Bruckmaier, 2014; Mulder & Gruber, 2011). Die vorliegende Arbeit verpflichtet sich dem Paradigma der professionellen Kompetenz, das den Blick auf personale Merkmale einer Lehrkraft richtet (Erklärkompetenz), die als Voraussetzung für die Berufsausübung (Erklären als Kerntätigkeit im Unterricht) angesehen werden. Dabei rücken jedoch eher Wissen und Überzeugungen von Lehrkräften als deren Charaktereigenschaften in den Vordergrund, wie es im Persönlichkeits-Paradigma gängig war (Krauss, 2020).

2.2.2 Professionelle Kompetenzen von Lehrkräften

Der kompetenztheoretische Ansatz führt das unterrichtliche Handeln einer Lehrkraft auf ihre dafür notwendigen Voraussetzungen, ihre *Kompetenzen*, zurück (Terhart, 2011). Das Paradigma wird gerade wegen der Fokussierung auf kognitive Kompetenzen (z. B. Wissen und Überzeugung von Lehrkräften) als Basis für professionelles Handeln eng mit dem Expertiseansatz in Verbindung gebracht (Krauss, 2020). Unterrichtliches Erklären kann in diesem Zusammenhang nicht nur als Teil fachdidaktischer Kompetenz, sondern auch als domänenspezifische Expertise einer Lehrkraft angesehen werden. Um Konzeptualisierungen fachübergreifender beziehungsweise musikbezogener Erklärkompetenz zu entfalten, ist im Folgenden der Begriff Kompetenz sowie das Paradigma professioneller Kompetenz zu erläutern. Davon ausgehend soll in einem weiteren Schritt dargelegt werden, welche Modelle professioneller Kompetenz von (Musik-)Lehrkräften vorliegen und inwieweit darin Aspekte über Erklären beziehungsweise das Erklären-Können modelliert sind. Abschließend werden Möglichkeiten der Operationalisierung von Erklärkompetenz anhand von Forschungsprojekten herausgearbeitet, in denen gemessen oder beurteilt wird, wie gut Lehrkräfte erklären können.

Zum Begriff Kompetenz

Kompetenz wird nach Klieme et al. (2003) definiert als „die bei Individuen verfügbaren oder von ihnen erlernbaren kognitiven Fähigkeiten und Fertigkeiten, bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (S. 72). Die Autorinnen und Autoren stützen sich dabei auf Überlegungen Weinerts (2001), der mit seinen theoretischen Überlegungen die grundsätzliche Idee verfolgte, den bis dahin in Bezug auf Schüler*innen verwendeten Begriff *Leistung* durch *Kompetenz* zu ersetzen (Klieme et al., 2003). Der Begriff Kompetenz wurde aus der Expertiseforschung übernommen und auf das schulische Umfeld übertragen, indem dadurch zunächst das Wissen und Handeln von Schüler*innen beschrieben werden sollte; er diente erst im weiteren Verlauf auch als Rahmung der professionellen Kompetenz von Lehrkräften (Krauss, 2020).

Der Kompetenzbegriff Weinerts (2001) liegt musikpädagogischen Forschungsprojekten wie KoMus (Jordan et al., 2012; Knigge, 2011; Niessen et al., 2008) oder KOPRA-M (Hasselhorn, 2015; Hasselhorn & Lehmann, 2014) zugrunde, die musikbezogene Kompetenzen von Schüler*innen erfassen. Die Übernahme dieses bildungswissenschaftlichen Kompetenzbegriffs wird in der Musikpädagogik jedoch intensiv diskutiert (Knigge, 2014). Kritisiert werden unter anderem der fehlende Bezug zwischen musikbezogener Kompetenz und musikalischer Bildung (Vogt, 2008) und die Fokussierung auf ein kognitiv geprägtes Verständnis von Kompetenz (Brenk, 2014). Nach Rolle (2008) sei „zu bezweifeln, dass sich das Ziel von Musikunterricht angemessen als Problemlösefähigkeit beschreiben lässt“ (S. 46). Um den vorwiegend kognitiven Aspekt des Problemlösens aufzubrechen, bezieht sich Knigge (2014) auf weitere Überlegungen Weinerts (2001),

wonach – und aus musikpädagogischer Sicht passgenauer – das Problemlösen grundsätzlich als „Bewältigen von Anforderungssituationen“ (Knigge, 2014, S. 114) zu verstehen ist. Darüber hinaus beobachtet der Autor auch, dass der Kompetenzbegriff mittlerweile in nahezu allen musikpädagogischen Kontexten präsent ist (Knigge 2014). So ist auch die professionelle Kompetenz von Musiklehrkräften Gegenstand des musikpädagogischen Forschungsinteresses geworden (siehe Puffer & Hofmann, 2017) und ist Grundlage für das vorliegende Forschungsvorhaben.

Das Paradigma professioneller Kompetenz

Der kompetenztheoretische Ansatz zeichnet sich unter anderem dadurch aus, dass nicht nur „Kompetenzbereiche und Wissensdimensionen“ zur Bewältigung von unterrichtlichen Anforderungssituationen definiert werden, sondern auch eine prinzipielle „Erlernbarkeit eines erfolgreichen Lehrerhandelns“ (Terhart, 2011, S. 207) anzunehmen ist. Kritik besteht allerdings darin, dass Forschungsvorhaben, die den kompetenztheoretischen Ansatz verfolgen, den Begriff Kompetenz häufig nicht vollumfänglich definieren und im Zuge dessen unter dem begrifflichen Deckmantel der Kompetenz lediglich Wissensbestände messen (Blömeke, König et al., 2015; Mulder & Gruber, 2011).

Als Reaktion auf diese Kritik werden Kompetenzmodelle modifiziert und erweitert, indem beispielsweise das professionelle Handeln – nach der Definition von Weinert (2001) wichtiger Bestandteil von Kompetenz – auch auf Modellebene sichtbar gemacht und in Beziehung zur Dispositionsebene gebracht wird (Blömeke, Gustafsson et al., 2015). Gleichzeitig dokumentiert sich dadurch die enge Verbindung zur Expertiseforschung, die das professionelle Wissen in einer Domäne als entscheidende Grundlage für Expertise und damit auch professionelles Handeln ansieht (Mulder & Gruber, 2011; Terhart, 2011). Die Unterschiede zwischen beiden Ansätzen, dass selbstregulative Fähigkeiten oder motivationale Aspekte zur professionellen Kompetenz von Lehrkräften gezählt werden (z. B. Kunter et al., 2011), jedoch für die Expertiseforschung eine untergeordnete Rolle spielen (Krauss, 2020), fallen bezüglich des Erkenntnisinteresses der vorliegenden Arbeit im weiteren Verlauf nicht ins Gewicht.

Modellierung

Die musikpädagogische Forschungslage zur Frage, wie Erklären als fachdidaktische Kompetenz einer Musiklehrkraft modelliert werden kann, erweist sich als defizitär (Frei, 2019). Zwar legten Puffer und Hofmann (2017) zunächst in Zusammenhang mit dem FALKO-M-Professionswissenstest auch eine Modellierung des Fachwissens und fachdidaktischen Wissens von Musiklehrkräften vor, die zu einem umfassenden Modell professioneller Kompetenzen von Musiklehrkräften erweitert wurde (Puffer, 2021). Hinsichtlich der Konzeptualisierung von Erklärkompetenz kann die Modellierung jedoch nur für eine erste Annäherung dienen. Denn hinsichtlich des Erklärens fällt einschränkend ins Gewicht, dass Puffer und Hofmann (2017) die Facette *Wissen über Erklären und*

Repräsentieren im Gegensatz zu allen anderen Fächern der FALKO-Forschungsgruppe (Krauss et al., 2017) anders bezeichnen und auf den Begriff *Erklären* verzichten. Gewählt wurde stattdessen die Bezeichnung „Wissen um musikbezogene Instruktionsstrategien“ (Puffer & Hofmann, 2017, S. 254) – das Wort *Instruktion* dient dabei als Sammelbegriff im Sinne von *instruction* aus dem angloamerikanischen Diskurs (z. B. Dorfman, 2013). Puffer und Hofmann (2017) zählen dazu beispielsweise das „Wissen um anschauliche Illustrationen und ‚anhörliche‘ Beispiele“ oder „Wissen um verschiedene Arten und Formen der Demonstration von Lerninhalten, einschließlich musikpraktischen ‚Vormachens‘“ (S. 254).

Unabhängig von der Funktion des Sammelbegriffs *Instruktion* bleibt das Verhältnis von Formen des Erklärens und Instruierens anhand der Handlungen einer Musiklehrkraft bei dieser Konzeptualisierung unklar. Eine Modellierung guten Erklärens und guten Instruierens wurde an anderer Stelle vorgeschlagen (Frei et al., 2022). Auf diese geht die vorliegende Arbeit in Kapitel 2.4.2 ausführlich ein. Inwieweit auf Basis dieses Modells von Erklärqualität auch die Erklärkompetenz von Lehrkräften modelliert und empirisch validiert werden kann, ist die Aufgabe weiterer Forschungsbemühungen. Ohne eine ausführliche Modellierung vorzunehmen, wird an anderer Stelle vorgeschlagen, von *musikbezogenen Strategien im Erklären und Instruieren* (Frei et al., i. Vorb.) zu sprechen und damit zur weiteren Konkretisierung der Kompetenzfacette beizutragen.

Aufgrund des Mangels an einer umfassenden musikbezogenen Konzeptualisierung von Erklärkompetenz sind deshalb fachfremde Konzepte zur (Erklär-)Kompetenz von Lehrkräften heranzuziehen (siehe dazu Puffer & Hofmann, 2016). Potenzielle Kandidaten hierfür sind beispielsweise Kompetenzmodelle aus anderen Fachdidaktiken (z. B. Baumert & Kunter, 2011), allgemeine Standards zur Ausbildung von Lehrkräften (z. B. Frey, 2014) oder Anforderungen an Studienabsolvent*innen aller Lehrämter (KMK, 2019). Die Bildungsadministration schlägt bundesländerübergreifend einheitliche Anforderungen hinsichtlich der Kompetenz angehender Lehrkräfte aller Fächer vor (KMK, 2019). Aus diesen können jedoch keine direkten Rückschlüsse auf die Erklärkompetenz von Musiklehrkräften gezogen werden, da Erklären weder im Fach Musik noch in den anderen Unterrichtsfächern – mit Ausnahme von Informatik – als Kompetenz einer Lehrkraft explizit genannt wird (KMK, 2019). Einen Überblick zu Standardmodellen der Lehrer*innenbildung gibt Frey (2014). In dieser Zusammenschau zeigt sich ebenfalls, dass Erklärkompetenz nicht explizit modelliert wird. Allerdings kann angenommen werden, dass die Kompetenz im Erklären beispielsweise unter Aspekten wie Methodenkompetenz mitgedacht wird und somit zumindest implizit in diesem Modell Berücksichtigung findet.

Weiterhelfen können jedoch Kompetenzmodelle, die das Wissen von Lehrkräften in verschiedene (Kompetenz-)Bereiche²⁸ unterteilen (z. B. Baumert & Kunter, 2011). Bevor der Fokus auf der Konzeptualisierung von Erklären in unterschiedlichen Modellen liegt,

28 In den Ausführungen zum Projekt COACTIV (Kunter et al., 2011) wird von Kompetenzbereichen gesprochen, wobei dann von Fachwissen statt Fachkompetenz oder Fachdidaktischem Wissen statt Fachdidaktischer Kompetenz gesprochen wird. Damit wird der Tatsache Rechnung getragen, dass in COACTIV ausschließlich Wissensaspekte gemessen wurden, diese jedoch als Teil von Kompetenz gedacht werden.

geht es zunächst darum, aufzuzeigen, wie Kompetenzmodelle entwickelt, modifiziert und teilweise in weiteren Schritten fachspezifisch adaptiert beziehungsweise ausgeschärft werden. Der folgenden Auswahl an Kompetenzmodellen liegt zugrunde, dass alle Modellierungen auf dem Kompetenzbegriff nach Weinert (2001) basieren und in das Kaskaden-Modell von Krauss et al. (2020) eingeordnet werden können. Dieses greift verschiedene Forschungsparadigma auf und integriert darüber hinaus unterschiedliche Konzeptualisierungen zur Kompetenz von Lehrkräften.

Das Kompetenzmodell der COACTIV-Studie (Kunter & Voss, 2011) rekurriert auf Shulmans fachübergreifende Taxonomie zum professionellen Wissen von Lehrkräften (1986, 1987). Kunter und Voss (2011) nehmen bezüglich der Modellierung professioneller Kompetenz von Mathematiklehrkräften Domänenspezifität an. Dabei folgen sie der Vorstellung, dass das Denken, Wissen und Handeln von Lehrkräften in hoher Abhängigkeit zum Fachinhalt steht (Kunter & Voss, 2011). Während die Autor*innen der COACTIV-Forschungsgruppe zunächst ein generisches Modell professioneller Kompetenz von Lehrkräften aufgestellt haben, erfolgte in einem zweiten Schritt die Spezifizierung hinsichtlich Mathematiklehrkräften (Baumert & Kunter, 2011). Die COACTIV-Studie kann als Ergebnis und Reaktion auf den *PISA-Schock* in Deutschland im Jahr 2000 verstanden werden. Sie sucht nach Gründen für den Mangel an mathematischen Kompetenzen deutscher Schüler*innen. Daher wurden in der COACTIV-Studie (Kunter et al., 2011), begleitend zum PISA-Test 2003, die Kompetenzen der jeweiligen Lehrkräfte gemessen, deren Schüler*innen an PISA beteiligt waren. Die fachliche Fokussierung der damaligen PISA-Tests auf mathematische Kompetenzen mag als ein Grund dafür gelten, dass die Mathematikdidaktik in Deutschland bezüglich der Modellierung und Messung von professioneller Kompetenz bei Lehrkräften eine gewisse Vorreiterrolle ausfüllt (Krauss et al., 2017) – weitere Large-Scale-Studien zu Kompetenzen von Mathematiklehrkräften beziehungsweise der Qualität von Mathematikunterricht wie MT21 (Blömeke et al., 2010a), TEDS-M (Blömeke et al., 2010b) oder die so genannte Pythagoras-Studie (Klieme et al., 2006) unterstreichen dies.

Zudem dienen die Modelle zur Kompetenz von Mathematiklehrkräften auch als Grundlage für darauf aufbauende Kompetenzmodellierungen anderer Fächer. Beispielsweise folgte auf die rein mathematik-spezifische TEDS-M-Studie (Blömeke et al., 2010b) das Forschungsprojekt TEDS-LT (Blömeke et al., 2011), dessen Forschungsinteresse neben Mathematik auf den Fächern Deutsch und Englisch liegt. Grundlegendes Ziel dieser internationalen Vergleichsstudie ist es, das professionelle Wissen von (angehenden) Mathematik-, Deutsch- und Englischlehrkräften zu untersuchen und dabei Erkenntnisse über die Wirksamkeit der Lehrkräfteausbildung zu gewinnen. Auch das FALKO-Projekt (Krauss et al., 2017) hat Professionswissenstests für sechs Unterrichtsfächer entwickelt und folgt dem methodischen Vorgehen der COACTIV-Studie (Kunter et al., 2011) beziehungsweise adaptiert Ausschnitte aus dessen Kompetenzmodell (Kunter & Voss, 2011), ohne dabei jedoch auf fachspezifische Modifikationen zu verzichten.

Grundlegendes Ziel der FALKO-Studie war die Entwicklung und Validierung von Professionswissenstests für die Bereiche Fachwissen und Fachdidaktisches Wissen der Fächer Deutsch, Englisch, evangelische Religion, Latein, Musik und Physik (Krauss et

al., 2017). Ein Ergebnis der Studie war, dass Fachwissen und Fachdidaktisches Wissen in den jeweiligen Kompetenzmodellierungen aller beteiligter Fächer – wie auch schon für das Fach Mathematik in der COACTIV-Studie (Krauss et al., 2011) – als zwei eigenständige, sich aber dennoch überschneidende Bereiche dargestellt werden können (Lindl & Krauss, 2017). In diesem Zusammenhang wurden diesbezüglich jedoch keine fachübergreifenden Erkenntnisse zur Facette Erklärungswissen präsentiert.

Ein anderes Modell, das Kompetenz nicht mehr rein auf der Wissensebene abbildet (z. B. Kunter et al., 2011), sondern als Kontinuum zwischen den Polen Disposition und Performanz versteht, entwickelten Blömeke, Gustafsson et al. (2015). Sie betrachten dabei das Unterrichtshandeln einer Lehrkraft als „Kompetenz im weiteren Sinne“ (Krauss, 2020, S. 155). Als Mittler zwischen der Dispositions- und der Performanzebene fungieren situationsspezifische Fähigkeiten und transformieren die Voraussetzungen einer Lehrkraft auf kognitiver sowie affektiv-motivationaler Ebene in konkretes Unterrichtshandeln (Blömeke, Gustafsson et al., 2015). Im Zuge dieser „situation-specific skills“ (Blömeke, Gustafsson et al., 2015, S. 7) wird das Konzept der Professionellen Wahrnehmungsfähigkeit als Aspekt von Lehrer*innenexpertise in die Diskussion eingebracht, das Goodwin (1994) und später Sherin (2001) unter dem Begriff *professional vision* als Facette der Expertise von Lehrkräften konzeptualisiert haben. Im deutschsprachigen Raum diente es zuerst bei Seidel et al. (2010) als theoretischer Rahmen für die Erfassung von Zielorientierung, Lernbegleitung und Lernatmosphäre beim Unterrichtshandeln von Lehrkräften. Als charakteristisch gelten dabei die aufmerksame und selektive Identifikation (*selective attention* und *noticing*) sowie eine daran anschließende theoriebasierte Interpretation (*reasoning*) einer Unterrichtssituation, um auf Basis dessen professionelles Handeln zu ermöglichen (Seidel et al, 2010; van Es & Sherin, 2008). Innerhalb der Musikpädagogik gewinnt das Konzept der Professionellen Wahrnehmung durch erste Forschungsvorhaben zunehmend an Bedeutung (z. B. Harnischmacher & Hofbauer, 2011; Lenord, 2020; Puffer, 2021), ohne dabei allerdings auf die Untersuchung der Erklärkompetenz von Musiklehrkräften Bezug zu nehmen.

Der Fokus musikpädagogischer Kompetenzforschung lag generell – dies gilt im Übrigen auch für den Bereich der allgemeinen Lehr-Lernforschung (Trautmann & Wischer, 2016) – weniger auf dem Lehren und den Lehrenden, sondern mehr auf der Untersuchung des Lernens und der Lernenden (Lehmann-Wermser & Krause-Benz, 2013, S. 8; siehe dazu z. B. Harnischmacher et al., 2019; Hasselhorn, 2015; Heß, 2011; Jordan et al., 2012; Knigge, 2011; Lill et al., 2019). Bezüglich der professionellen Kompetenz von Musiklehrkräften lässt sich jedoch in den letzten Jahren ein verstärktes Forschungsinteresse beobachten (z. B. Dreßler, 2017; Fiedler & Müllensiefen, 2016; Göllner & Niessen, 2016; Harnischmacher et al., 2017; Niessen, 2013; Puffer, 2021). In diesem Zusammenhang rekurren auch einige Studien auf die Wissenstaxonomie Shulmans (1986) und leisten einen Beitrag zur Modellierung professioneller Kompetenzen von Musiklehrkräften (z. B. Godau & Fiedler, 2018; Puffer & Hofmann, 2017), nachdem diese Taxonomie in der musikpädagogischen Forschung lange keine Rolle zu spielen schien (Hofmann, 2011).

Einen für die Musikpädagogik ersten umfassenden Vorschlag für die Modellierung professioneller Kompetenz von Musiklehrkräften legte Puffer (2021) mit einem erweiterten Modell professioneller Kompetenzen von Musiklehrkräften vor (siehe auch Puffer & Hofmann, 2022). Dabei werden einige Ansätze in die Modellierung mit einbezogen, die in den vorherigen Absätzen bereits vorgestellt wurden und den bildungswissenschaftlichen Diskurs um professionelle Kompetenz von Lehrkräften in den letzten Jahren geprägt haben (siehe Blömeke, Gustafsson et al., 2015; Kunter et al., 2011). Basierend auf diesen Überlegungen integriert Puffer (2021) „musikpraktisch-künstlerische Fähigkeiten und Fertigkeiten als wesentlichen Bestandteil fachspezifischer professioneller Kompetenz“ (S. 48), die sich aus kognitiven, affektiv-motivationalen, perzeptiven und sensomotorischen Komponenten zusammensetzt. Die vorliegende Arbeit orientiert sich an dieser Modellierung professioneller Kompetenz von Musiklehrkräften und fokussiert unterrichtliches Erklären als Aspekt des fachdidaktischen Wissens (FDW) beziehungsweise fachdidaktischer Kompetenz.²⁹ In der Darstellung werden jedoch zunächst die mit ihm in direkter Verbindung stehenden Elemente Fachwissen (FW), (Allgemein-)Pädagogisches Wissen (PW) sowie die berufsfeldspezifischen musikpraktisch-künstlerischen Fertigkeiten (BF) von fachdidaktischem Wissen abgegrenzt.

Pädagogisch-psychologisches Wissen (Voss & Kunter, 2011) beziehungsweise die pädagogische Kompetenz von Lehrkräften umfasst Bereiche außerhalb des Unterrichts, wie zum Beispiel die Beratung von Eltern wie Schüler*innen oder Aufgaben der Schulentwicklung, und Bereiche innerhalb des Unterrichts, wie beispielsweise Klassenführung (siehe Mulder et al., 2017). Da eine ausführliche Darstellung dieses Kompetenzbereichs in der vorliegenden Arbeit nicht möglich ist, wird auf weiterführende Literatur verwiesen (z. B. Mulder et al., 2017; Voss et al., 2015). Dennoch sollen zumindest ausgewählte Komponenten pädagogischer Kompetenz von Lehrkräften skizziert werden, die für das Unterrichten im Fach Musik relevant erscheinen und mit dem Erklären als Unterrichtshandlung einer Lehrkraft in Verbindung stehen können.

Eberhard (2010) konnte zeigen, dass es aus Sicht von Lehrkräften wie Schüler*innen im Musikunterricht tendenziell häufiger zu Unterrichtsstörungen als in anderen Fächern

29 Zu weiteren Aspekten professioneller Kompetenzen von Musiklehrkräften, wie selbstregulative oder affektiv-motivationale Fähigkeiten sowie Überzeugungen, liegen in der Musikpädagogik verschiedene Befunde vor: Busch und Kranefeld (2013) fragten danach, inwieweit Überzeugungen, Motivation, Selbstregulation und Erfahrung einen Einfluss auf die Förderung nach individuellen Lernvoraussetzungen von Schüler*innen haben und konnten – unter der Limitation einer kleinen Stichprobe – erste Anhaltspunkte liefern, dass vor allem Selbstregulation, gemessen an der Selbstwirksamkeitserwartung der Lehrkräfte, als prädiktiv valide für die Förderung individueller Lernvoraussetzungen angesehen werden kann. Auch die Forscher*innengruppe um Harnischmacher widmet sich Aspekten professioneller Kompetenz von Musiklehrkräften. Hofbauer (2017) konnte zeigen, dass Motivation und Expertise von Musiklehrkräften einen Einfluss auf selbstregulative Fähigkeiten wie Strategien zur Stressbewältigung im Musikunterricht nehmen. Harnischmacher et al. (2017) untersuchten außerdem Überzeugungen von Musiklehrkräften bezüglich eines kompetenzorientierten und damit subjektorientierten Musikunterrichts. Die Autor*innen konnten dabei unter anderem ihre Hypothese bestätigen, dass „Kompetenzorientierung durch Motivation musikpädagogischen Handelns, Selbstreflexion und Musizierpraxis vorhergesagt werden kann“ und „Motivation als Mediator für Lehrerfahrung und Selbstreflexion“ (Harnischmacher et al., 2017, S. 282) fungiert.

kommt. Dies wird unter anderem am geringen Stellenwert des Faches Musik oder an fachspezifischen Situationen wie gemeinsamem Musizieren mit Instrumenten festgemacht, da diese einen besonderen Aufforderungscharakter zur (unsachgemäßen) Bedienung aufweisen (Eberhard, 2010). Vor diesem Hintergrund fallen für das gemeinsame Musizieren – nun im Kontext einer Untersuchung von Grundschulmusikunterricht – Begriffe wie Reibungslosigkeit und Schwung, deren Gewährleistung als zentrale Anforderung an Musiklehrkräfte gestellt wird (Niessen, 2013). In Bezug auf die Unterrichtsgestaltung und die dabei angewendeten Lehr- und Lernkonzepte weisen Musikschullehrkräfte, die vorwiegend über Erfahrung im Einzel- beziehungsweise Kleingruppenunterricht verfügen, beim Unterrichten einer Schulklasse im Musikunterricht auf den Status von Noviz*innen zurück (Kranefeld, 2015). Dies zeigt sich sowohl im Umgang mit Heterogenität (Göllner & Niessen, 2016; siehe dazu auch Jank, 2013), die in einer Schulklasse häufig mehr gegeben ist als in einer Kleingruppe – ganz abgesehen von Einzelunterricht –, als auch bei Inszenierungstechniken hinsichtlich der Erstbegegnung von Schüler*innen mit Instrumenten, deren spezifischer Klangerzeugung und Funktionsweise (Linn, 2017; Niessen, 2013).

Welche wichtige Rolle Kompetenzen in der Klassenführung für angehende Musiklehrkräfte spielen, konnte Puffer (2013) in einer qualitativen Untersuchung zeigen, da in Nachbesprechungen zu Praktikumsstunden von den beteiligten Studierenden vor allem Aspekte zur Prävention von oder zum Umgang mit Unterrichtsstörung thematisiert wurden. Die Autorin kommt an anderer Stelle zu dem Schluss, dass Musiklehrkräfte über domänenspezifisches Wissen von Klassenführung verfügen sollten, weshalb sie diese eher als fachdidaktische denn als allgemeinpädagogische Wissensfacette konzeptualisiert (Puffer, 2021).

Das Fachwissen nimmt im Diskurs um die professionelle Kompetenz von Lehrkräften eine wichtige Rolle ein. So stellt Neuweg (2014) – zumindest in Bezug auf Mathematiklehrkräfte – weitgehende Einigkeit darin fest, dass Fachwissen eine notwendige, jedoch keine hinreichende Bedeutung für die Qualität von Unterricht einnimmt (Baumert & Kunter, 2006). Innerhalb der COACTIV-Studie kategorisierten Krauss et al. (2011) das Fachwissen von Mathematiklehrkräften in vier Ebenen: (1) Alltagswissen, (2) Schulstoffwissen „auf dem Niveau eines durchschnittlichen bis guten Schülers“ (S. 142), (3) tieferes Verständnis des Schulstoffs und (4) reines Universitätswissen, das vom Schulstoff losgelöst ist (Neuweg, 2014). In Bezug auf gutes Erklären kann hohes Fachwissen von Lehrkräften unterschiedlichen Einfluss nehmen. Während es einerseits eine Voraussetzung für eine fachlich korrekte Erklärung darstellt (Neuweg, 2014), kann ausgeprägtes Expert*innenwissen andererseits auch dazu führen, die Erklärung nicht mehr an den (Vor-)Wissensstand einer Adressat*innengruppe anpassen zu können – dieser Effekt wird als *expert blind spot* beschrieben (siehe dazu z. B. Renkl et al., 2006).

Eine erste empirisch validierte Modellierung des Fachwissens und des fachdidaktischen Wissens von Musiklehrkräften entwickelten Puffer und Hofmann (2017) im Rahmen des FALKO-M-Projekts „analog zum Vorgehen des COACTIV-Forschungsteams und unter Berücksichtigung fachspezifischer Anforderungen“ (S. 253). Mit dem Ziel, das Wissen von Musiklehrkräften auf Stufe drei und damit „auf einem möglichst

„unterrichtsnahen‘ Niveau zu erheben“ (Puffer & Hofmann, 2016, S. 110; siehe dazu auch Krauss et al., 2011), sowie bezugnehmend auf Forderungen der Bildungsadministration subsumieren die Autorin und der Autor unter Fachwissen von Musiklehrkräften „Wissen über Musik unterschiedlicher Epochen, Kulturen und Genres sowie über historische, kulturelle und soziale Kontexte, Wissen über Musiker, Wissen über Notation und Beschreibung von Musik sowie Wissen um Musikmedien und -technik“ (Puffer & Hofmann, 2017, S. 257).

Neben der erfolgreichen Validierung des FALKO-M-Professionswissenstests waren zentrale Ergebnisse, dass hohes Fachwissen von Musiklehrkräften eine Voraussetzung für ausgeprägtes fachdidaktisches Wissen darstellt und sich hinsichtlich der Schulform (Gymnasium – Nicht-Gymnasium) beziehungsweise des Status (Studierende – Lehrkräfte) unterscheidet (Puffer & Hofmann, 2017). Auffällig war außerdem, dass vor allem angehende Lehrkräfte nicht gymnasialer Lehramtsstudiengänge über deutlich geringeres Fachwissen als Studierende des gymnasialen Lehramts verfügen, während der Unterschied zwischen den Lehrkräften beider Gruppen im Durchschnitt nicht ins Gewicht fiel, wenngleich die nicht gymnasialen Lehrkräfte im Bereich Rock-Pop leichte Vorteile gegenüber den gymnasialen Lehrkräften zeigten (Puffer & Hofmann, 2017).

Ergänzend zu diesen zwei fachübergreifend gültigen Wissensfacetten integriert Puffer (2021) in ihrer Modellierung auch Komponenten des Professionswissens, die ausschließlich für das Unterrichten einer Musiklehrkraft relevant sind. Im Gegensatz zu *allgemeinen* musikpraktisch-künstlerischen Fertigkeiten, die in Bezug zum Fachwissen verortet werden und „über die beispielsweise auch Professionelle aus anderen Musikberufen oder besonders fortgeschrittene musikalische Laien verfügen können“ (Puffer, 2021, S. 49), werden unter der *berufsfeldspezifischen* Kategorie beispielsweise musikpraktisch-künstlerische Fertigkeiten im Schulpraktischen Klavierspiel subsumiert – bereits der FALKO-M-Test erhebt dies dahingehend, dass die Teilnehmenden eine Liedbegleitung für ein Begleitinstrument ausnotieren sollten (Puffer & Hofmann, 2017). In der konkreten Handlungssituation des Unterrichts zählen zu professionellen Fertigkeiten im Schulpraktischen Klavierspiel beispielsweise, mit Hilfe des Begleitinstruments „musikalische Sachverhalte ‚anhörlich‘ zu machen“, oder beim Singen und Musizieren „im Klassenverband je nach Lernstand variabel zu begleiten“ (Puffer, 2021, S. 49). Aus diesen Gründen werden die berufsfeldspezifischen musikpraktisch-künstlerischen Fertigkeiten mit dem Bereich fachdidaktisches Wissen verknüpft, der nun zunächst wieder aus einer fachspezifischen Perspektive beleuchtet wird und für die vorliegende Arbeit zentral ist. Zur Frage, inwieweit diese Verbindungslinien auch für die Konzeptualisierung von Erklären im Musikunterricht relevant sind, geht die vorliegende Arbeit in Kapitel 2.4.1 ausführlicher ein.

Das fachdidaktische Wissen von Lehrkräften ist dafür verantwortlich, Fachinhalte didaktisch aufzubereiten und ist für das Fach Mathematik in Bezug auf den Leistungsfortschritt von Schüler*innen prädiktiv valide (Krauss et al., 2011). Die Konzeptualisierung des fachdidaktischen Wissens in der COACTIV-Studie basiert auf drei Eckpunkten: Dem (1) Zugänglichmachen von (2) Inhalten für (3) Schüler*innen (Krauss et al., 2011). Wenngleich diese drei Aspekte inhaltlich nicht unabhängig voneinander zu verstehen sind, führen sie doch zu drei zentralen Facetten fachdidaktischen Wissens, die in der

COACTIV-Studie als „Wissen über das multiple Lösungspotenzial von Mathematikaufgaben“, „Wissen über typische Schülerfehler und -schwierigkeiten“ und „Wissen über Erklären und Repräsentieren“ (Krauss et al., 2011, S. 138–139) konzeptualisiert wurden. Im FALKO-Projekt (Krauss et al., 2017) wurde in Anlehnung an die COACTIV-Studie nicht nur Fachwissen, sondern auch fachdidaktisches Wissen von Lehrkräften der sechs beteiligten Unterrichtsfächer modelliert. Zwar wurden an einzelnen Stellen fachspezifische Adaptionen notwendig, jedoch konnten zwei Facetten über alle Unterrichtsfächer hinweg weitestgehend gleich konzeptualisiert werden: Das Wissen über Kognitionen von Schüler*innen und das Erklärungswissen (Krauss et al., 2017).³⁰

Für die Musikpädagogik konzeptualisieren Puffer und Hofmann (2017) das fachdidaktische Wissen von Musiklehrkräften als „Wissen über musikbezogene Kognitionen, Lernwege und Schwierigkeiten von Schüler*innen“, als „Wissen über das Potential von Materialien für musikbezogene Lehr-/Lernprozesse“ und als „Wissen über musikbezogene Instruktionsstrategien“ (Puffer & Hofmann, 2017, S. 253–254). Hinsichtlich des Wissens um typische Fehler und mögliche Missverständnisse von Schüler*innen beziehen sie sowohl den kognitiven als auch den musikspezifisch relevanten psychomotorischen Bereich mit ein: Musiklehrkräfte sollten etwa beim gemeinsamen Musizieren sowohl in der Lage sein, *Brummer* hörenderweise zu erkennen, als auch „potenzielle Probleme beim Erlernen eines Mitspielsatzes zu antizipieren“ (Puffer & Hofmann, 2017, S. 254–255). Die Hinzunahme des psychomotorischen Bereichs stellt eine fachspezifische Adaption der ursprünglichen Modellierungen dar (siehe Krauss et al., 2011; Shulman, 1986).

Die zweite Facette wurde „in Anlehnung an das Konzept eines *curriculum knowledge* modelliert“ (Puffer, 2021, S. 32; siehe auch Puffer & Hofmann, 2016) und als „Wissen über das Potenzial von Materialien für musikbezogene Lehr-/Lernprozesse“ (Puffer & Hofmann, 2017, S. 255) ausgewiesen. Materialien wie Musikstücke oder Lieder, die im Musikunterricht musiziert beziehungsweise gesungen werden, bringen Potenziale mit sich; diese zu erkennen und didaktische Überlegungen miteinzubeziehen, sei dabei Aufgabe von Musiklehrkräften (Puffer & Hofmann, 2017). Vor dem Hintergrund, dass Schulbücher für das Unterrichten im Fach Musik mehr als Ideensteinbruch und weniger als Leitmedium dienen (Jünger, 2006), erfordert gerade die gezielte Auswahl von Aufgaben und Materialien eine hohe Fähigkeit, „Materialien zu beurteilen“ (Puffer & Hofmann, 2017, S. 255). Die Öffnung vom ursprünglichen Potenzial von Aufgaben (siehe Krauss et al., 2011) hin zu den in FALKO-M beschriebenen Potenzial von Materialien (Puffer & Hofmann, 2017) ist ebenfalls als fachspezifische Adaption zu werten. Mit Blick auf das Erklären im Musikunterricht ist beispielsweise relevant, anhand welcher Melodien, Rhythmen oder Lieder musiktheoretische Inhalte erklärt werden und ob diese Materialien

30 Die dritte Facette, die im COACTIV-Modell (Baumert & Kunter, 2011) verwendet wird, lautet: Wissen über das Potenzial von Aufgaben. Diese Facette wurde beispielsweise im Fach Physik durch das Wissen über Messen und Experimentieren (Schödl & Göhring, 2017, S. 216) ersetzt. Eine Erweiterung des Wissens um Potenziale von Aufgaben(-stellungen) wurde in den Fächern Latein (Lindl & Kloiber, 2017, S. 161) und Musik (Puffer & Hofmann, 2017, S. 255) vorgenommen.

angemessen ausgewählt sind (zu materialen Aspekten von Musikunterricht siehe z. B. Godau, 2018; Kranefeld, 2021).

Erklärkompetenz

Die dritte und für die vorliegende Arbeit besonders relevante Facette fachdidaktischen Wissens bezieht sich auf das Erklären und Repräsentieren³¹ von Inhalten und beschreibt die Fähigkeit, Schüler*innen Inhalte zugänglich beziehungsweise verständlich zu machen (Krauss et al., 2011). Sie steht mit den beiden eben genannten fachdidaktischen Facetten in engem Zusammenhang, da eine Lehrkraft beim Erklären nicht nur musikbezogene Lernwege der Schüler*innen berücksichtigen, sondern auch dabei auftretende Schwierigkeiten beheben kann. Diesbezüglich zeigen Ergebnisse aus COACTIV beispielsweise, dass ein hohes Wissen über Erklären und Repräsentieren mit einem hohen Wissen um die Kognitionen von Schüler*innen korreliert – nach Krauss et al. (2011) ein erwartungskonformer Befund, da „eine gute mathematische Erklärung neben Fachwissen auch Wissen über diesbezügliche Schülerkognitionen“ erfordert (S. 150). Erklären wird jedoch nicht nur als relevante professionelle Kompetenz einer Mathematiklehrkraft angesehen, sondern findet, wie bereits dargelegt, sowohl in vielen Professionswissenstests als auch Fragebogenerhebungen mit Schüler*innen zur Beschreibung von Unterrichtsqualität Berücksichtigung.

Im FALKO-Projekt (Krauss et al., 2017) konzeptualisieren alle sechs beteiligten Unterrichtsfächer die Facette Erklären und Repräsentieren, wobei fachliche Adaptionen bereits in den Bezeichnungen der Facette des Erklärungswissens deutlich werden. Während diese analog zur COACTIV-Studie als „Erklären und Repräsentieren“ in den Fächern Physik (Schödl & Göhring, 2017, S. 216), Deutsch (Pissarek & Schilcher, 2017) und Latein (Lindl & Kloiber, 2017) benannt wurde, gab es in den anderen Fächern bereits (leichte) Modifizierungen in den Bezeichnungen. Aufgrund der fehlenden Begründung dieses Vorgehens kann nur angenommen werden, dass damit fachspezifischen Besonderheiten Rechnung getragen werden sollte. Ergänzt wurde die Bezeichnung im Fach Englisch um die Perspektive der Schüler*innen, denn dort ist von „Wissen über schülergerechtes Erklären und Repräsentieren“ (Kirchhoff, 2017, S. 120) die Rede. Demgegenüber wird im Fach Evangelische Religion nur „Erklären“ (Fricke, 2017, S. 304) verwendet.

Im Fach Musik nehmen Puffer und Hofmann (2017), wie bereits dargelegt, auf theoretischer Ebene Anpassungen vor, indem sie die Wissensfacette weiter fassen und als „Wissen über musikbezogene Instruktionsstrategien“ (S. 254) ausweisen. Bei der Konzeptualisierung tragen sie der musizierend-praktischen Fachspezifität insofern Rechnung, als musikpraktisches Vormachen explizit zu Arten und Formen der Demonstration von Lerninhalten gezählt wird oder anschauliche Illustrationen um ‚anhörliche‘ Beispiele ergänzt werden (Puffer & Hofmann, 2017). Der dritte Aspekt bezieht sich auf musikbezogene Gestaltungs- und Übeprozesse mit einer Gruppe sowie auf das Wissen einer

31 Repräsentieren meint hier unterschiedliche Formen der Darstellung, beispielsweise als Text oder als Abbildung (siehe dazu z. B. Neubrand et al., 2011).

Musiklehrkraft, diese „zu initiieren, zu fördern und anzuleiten“ (Puffer & Hofmann, 2017, S. 254). Auf Basis dieser Modellierung wird deutlich, dass für die Vermittlung musikbezogener Inhalte akustische Elemente zentral sind (*musikpraktisches Vormachen, anhörliche Beispiele*) und daneben optische, taktile oder auch kinästhetische Aspekte eine Rolle spielen können (siehe dazu z. B. Laufer, 2018). Diese werden in Kapitel 2.4.2 ausführlich entfaltet.

Zusammenfassend berücksichtigen die bisher genannten fachspezifischen wie fachübergreifenden Modellierungen die Rolle des Erklärens unterschiedlich. Das Modell der COACTIV-Studie zählt das Erklärungswissen explizit zu einer der drei zentralen fachdidaktischen Kompetenzfacetten (Baumert & Kunter, 2011). Außerdem integrieren Blömeke et al. (2008) das Erklären in mathematikspezifische Studien wie MT21 sowie deren Nachfolgestudie TEDS-M (Blömeke et al., 2010). Während Erklären auch im mathematischen Teil von TEDS-LT eine große Rolle spielt (Buchholz et al., 2011), kann es in der Kompetenzmodellierung professionellen Wissens von Deutsch- und Englischlehrkräften in TEDS-LT, wenn überhaupt, nur als Randbedingung wahrgenommen werden (Bremrich-Vos et al., 2011). Hingegen modellieren die sechs Unterrichtsfächer der FALKO-Studie (Krauss et al., 2017) das Wissen um Erklären und Repräsentieren – also auch in Deutsch (Pissarek & Schilcher, 2017) und Englisch (Kirchhoff, 2017) – als wichtige Facette des fachdidaktischen Professionswissens, wobei Musik, wie bereits dargelegt, eine Ausnahme macht. Innerhalb des bildungswissenschaftlichen Diskurses finden sich also Modelle, in denen Erklären explizit (Krauss et al., 2017; Kunter et al., 2011) als auch implizit (Blömeke et al., 2010b; Blömeke et al., 2011) als Facette fachdidaktischer Kompetenz einer Lehrkraft konzeptualisiert wird. Im Folgenden richtet sich ein vertiefter Blick auf Mess- und Beurteilungsverfahren zur Erklärkompetenz von Lehrkräften.

Messverfahren

Die Darstellung der Messverfahren erfolgt anhand des Kompetenzmodells von Blömeke, Gustafsson et al. (2015), da dieses in Bezug auf die professionelle Kompetenz von Lehrkräften sowohl eine Wissens- als auch eine Handlungsebene integriert und sich daher für Erklärkompetenz besonders gut eignet. In einem Teil der Untersuchungen wird das Professionswissen von Lehrkräften anhand von Tests gemessen und anhand dessen auf deren professionelle Kompetenz geschlossen. Beispielsweise erheben die Professionswissenstests der TEDS-M-Studie (Blömeke et al., 2010b), der COACTIV-Studie (Kunter et al., 2011) und der FALKO-Studie (Krauss et al., 2017) die Dispositionsebene von Lehrkräften. Sie arbeiten dabei mit Items, in denen skizzenhaft Unterrichtsszenarien beschrieben werden, denen gegenüber sich die Studienteilnehmer*innen „didaktisch professionell verhalten“ sollen (Krauss et al., 2011, S. 138).³² Neben Multiple-Choice-Aufgaben (z. B. FALKO-D³³, Pissarek & Schilcher, 2017) verwenden die Tests der FALKO-Gruppe vor allem offene Antwortformate, wie die Beispielitems *Rondo* aus FALKO-M (Puffer &

32 Für Item-Beispiele aus COACTIV siehe Krauss et al., 2011, S. 140–141.

33 FALKO-D ist der Professionswissenstest im Fach Deutsch.

Hofmann, 2017) oder *Gerundium und Gerundivum* aus FALKO-L³⁴ (Lindl & Kloiber, 2017) zeigen. Dabei erfolgt zunächst eine kurze Beschreibung der Unterrichtssituation (z. B. FALKO-E³⁵, Kirchoff, 2017) oder es werden akustische Stimuli, wie die Audioaufnahme einer singenden Klasse im Musikunterricht, präsentiert (Puffer & Hofmann, 2017).

Während der COACTIV-Test das Wissen über Erklären und Verständlichmachen mit zwölf der insgesamt 21 fachdidaktischen Items misst (Krauss et al., 2011), stellt der FALKO-M-Test fünf der elf Aufgaben aus dem Bereich des fachdidaktischen Wissens zu musikbezogenen Instruktionsstrategien (Puffer & Hofmann, 2017). Dem Vorgehen der COACTIV-Studie (Kunter et al., 2011) folgend thematisiert die FALKO-Gruppe bei den Items zum Erklären die Vermittlung von Inhalten sowohl explizit wie auch implizit in der Aufgabenstellung (Krauss et al., 2017). Als Beispiel für eine explizite Erwähnung werden im FALKO-M-Test für Musiklehrkräfte die Studienteilnehmer*innen in der Aufgabe *Rondo* nach ihrem Vorgehen im Unterricht gefragt, „um Schüler*innen der 6. Jahrgangsstufe die Rondo-Form zu vermitteln“ (Puffer & Hofmann, 2017, S. 259). Anschließend erfolgt die Bitte um kurze Skizzierung der dabei bevorzugten Inhalte und Methoden (Puffer & Hofmann, 2017).

Kritisiert werden bei diesem Verfahren vor allem zwei Aspekte: Zum einen bestehen Zweifel daran, ob Aussagen über die Kompetenz von Lehrkräften auf Basis der reinen Messung von Wissensbeständen getroffen werden können, da bei der Transformation von Wissen in unterrichtliches Handeln eine situationsbedingte Abhängigkeit vermutet wird, die Wissenstests nicht erfassen können (Blömeke, König et al., 2015). Zum anderen werde in den zugrunde liegenden Messmodellen die „Komplexität und Vielfalt des Lehrberufs“ (Mulder & Gruber, 2011, S. 433) nicht adäquat aufgegriffen. Um dieser Kritik zu begegnen und komplexe Handlungssituationen in den Blick zu nehmen, erheben weitere Studien die situationsspezifischen Erklärkompetenzen von Lehrkräften.

Schopf und Zwischenbrugger (2015) erhoben situationsspezifische Kompetenz im Erklären von Studierenden der Wirtschaftspädagogik, indem sie diese unter Vorgabe des Themas sowie der Adressat*innengruppe zur Anfertigung schriftlicher Erklärungen aufforderten, die geschulte Rater*innen in einem anschließenden Verfahren bewertet haben. Umfassender untersuchte Findeisen (2017) ebenfalls Studierende der Wirtschaftspädagogik und fokussierte dabei deren Erklärfähigkeit im Bereich Rechnungswesen. Die Autorin erfasste dabei nicht nur das Wissen der Studienteilnehmer*innen mittels eines Tests, sondern untersuchte zudem deren Erklärfähigkeit handlungsbasiert (Findeisen, 2017). Um performative Aspekte des Erklärens erheben zu können, wurden simulierte Erklärsituationen konstruiert, in denen Lehramtsstudierende einem „standardisierten Schüler“ (Findeisen, 2017, S. 155), der als Schauspieler fungierte, Sachverhalte des Rechnungswesens zur Vorbereitung auf eine potenzielle Prüfung erklärt haben. Dabei zeigte sich unter anderem, dass nur wenige Lehramtsstudierende über multiple Erklärungsansätze zu einer Thematik verfügen und es innerhalb der Stichprobe zu einer großen Varianz hinsichtlich der Erklärfähigkeit kommt (Findeisen, 2017). Operationalisiert wurde diese mittels

34 FALKO-L ist der Professionswissenstest im Fach Latein.

35 FALKO-E ist der Professionswissenstest im Fach Englisch.

Qualitätsaspekten guten Erklärens (z. B. fachlicher Gehalt oder Lernendenzentrierung), deren Bewertung von geschulten Ratern in Form von Ratingskalen zu den videografierten Erklärsituationen abgegeben wurde (Findeisen, 2017). Ähnlich wie in der eben vorgestellten Studie nutzen beispielsweise auch Kulgemeyer und Tomczyszyn (2015) simulierte Unterrichtssituationen, um die Fähigkeit im Erklären von angehenden Physiklehrkräften messen zu können. Das Setting besteht darin, dass die Proband*innen einer Schülerin oder einem Schüler physikalische Sachverhalte erklären und dieser Dialog videografiert sowie anschließend ausgewertet wird. Die Auswertung basiert unter anderem auf Qualitätskriterien guten Erklärens, die an anderer Stelle entwickelt wurden (Kulgemeyer & Schecker, 2013).

Andere Testverfahren, die der Kritik an *paper-&-pencil*-Tests begegnen und situationsspezifischere Kontextualisierungen in Erhebungen ermöglichen, greifen auf Videovignetten in der Funktion eines Stimulus zurück (Blömeke, Gustafsson et al., 2015; Rehm & Bölsterli, 2014; Seidel & Thiel, 2017), die als zumeist „kurze Szenen“ definiert werden und „kritische Probleme aufzeigen“ (Rehm & Bölsterli, 2014, S. 215). Dabei werden „reale, verdichtete Unterrichtsvideographien oder fiktionale, realitätsnahe Szenarien“ (Seifried & Wuttke, 2017, S. 306) dargestellt, die Antworten auf diese Szenarien hervorrufen möchten. Erhoben werden diese Antworten sowohl durch offene (z. B. Bruckmaier et al., 2016) als auch geschlossene Antwortformate (z. B. Tepner & Dollny, 2014).

Für eine inhaltliche Validierung hinsichtlich der Authentizität und fachlichen Korrektheit von Videovignetten empfehlen Ollesch et al. (2018), Urteile von Expert*innen einzuholen. Im Rahmen eines Testverfahrens können Ratings von Expert*innen außerdem als Bezugsnorm für die abgegebenen Antworten der Studienteilnehmer*innen dienen (Seidel et al., 2010). In diesem Zusammenhang entwickelten Tepner und Dollny (2014) ein Verfahren, das sogenannte aggregierte Expertenurteile in Verbindung mit Videovignetten verwendet. Der FALKO-M-Professionswissenstest (Puffer & Hofmann, 2017) verwendet dieses Verfahren, um die Testantworten der Studienteilnehmer*innen mit den Urteilen von Expert*innen in Beziehung setzen zu können. Hinsichtlich des Inhalts von Videovignetten ist zwischen Aufnahmen regulären Unterrichts und nachgestellten Unterrichtssituationen zu differenzieren. Sogenannte *scripted videos* stellen unter Verwendung eines Drehbuchs unterrichtsnahe Situationen nach (Piwowar et al., 2018). Dieses Vorgehen minimiert nicht nur datenschutzrechtliche Probleme, sondern bietet darüber hinaus auch die Möglichkeit, aus didaktisch relevanter Sicht Veränderungen gezielt vorzunehmen und Situationen zu kreieren, die dem Forschungsinteresse dienlich sind (Piwowar et al., 2018).

Im Allgemeinen wird dem Einsatz von Videovignetten gegenüber schriftlich skizzierten Unterrichtssituationen eine höhere ökologische Validität zugesprochen, weshalb der hohe Aufwand der Videoproduktion sowie der Erhebung in Online-Surveys in jedem Fall gerechtfertigt ist (Seidel & Thiel, 2017). Allerdings ist nach wie vor mit Einschränkungen der ökologischen Validität zu rechnen, da die Erhebungen im Kontext von künstlich hergestellten Laborsituationen durchgeführt werden (Döring & Bortz, 2016, S. 106). Bezüglich der Messung der Erklärkompetenz von Musiklehrkräften zeigt sich der musikpädagogische Forschungsstand abgesehen vom FALKO-M-Professionswissenstest (Puffer &

Hofmann, 2017) defizitär. Dementsprechend gibt es auch keine Erfahrungen hinsichtlich des Einsatzes von Videovignetten in diesem Kontext. Erste Bemühungen aus FALKO-M deuten jedoch an, dass das Erklären in musikbezogenen Aufgabenstellungen sowohl explizit als auch implizit thematisiert werden kann.

Beurteilungsverfahren

Um Rückschlüsse auf die (Erklär-)Kompetenz von Lehrkräften zu erhalten, dienen auch Beurteilungen von Schüler*innen, Studierenden oder Expert*innen wie erfahrenen Lehrkräften oder Fachdidaktiker*innen (z. B. DESI-Konsortium, 2008; Reusser & Pauli, 2003; Schopf & Zwischenbrugger, 2015; Wragg & Wood, 1984). Diese erfolgen allerdings auf zwei verschiedenen Ebenen: Einerseits wird in Form offen gestellter Fragen gebeten, Einschätzungen zu grundsätzlichen wichtigen Eigenschaften und Fähigkeiten einer guten Lehrkraft abzugeben. Andererseits werden Schüler*innen konkret aufgefordert, das Handeln und Verhalten ihrer Lehrkraft anhand vorgegebener Items zu bewerten.

Wragg und Wood (1984) legten Schüler*innen im Alter von 12 bis 16 Jahren 32 Aussagen über ein mögliches Verhalten und Handeln ihrer *Wunsch*-Lehrkraft vor. Neben den Eigenschaften, dass die Lehrperson beispielsweise gut zuhören oder schwächere Schüler*innen gut mitnehmen kann, erhielt die Fähigkeit, Inhalte klar erklären zu können, die höchste Zustimmung („The teacher would explain things clearly“, Wragg & Wood, 1984, S. 83). Damit knüpfen die Autoren an eine frühere Studie an, die zwar nur mit sieben Statements zum Handeln und Verhalten von Lehrkräften, jedoch mit einer Stichprobe von über 8.000 befragten Schüler*innen durchgeführt wurde (Hollis, 1935, zitiert nach Wragg & Wood, 1984). Auch diese Studie kam zu dem Ergebnis, dass Schüler*innen das geduldige Erklären einer Lehrkraft ebenfalls als deren wichtigste Fähigkeit wertschätzen (Wragg & Wood, 1984). Des Weiteren nennen Schüler*innen im Rahmen einer Videostudie zu Erklärsituationen im Mathematikunterricht (Wörn, 2014) auf die Frage nach bis zu drei Eigenschaften einer guten Mathematiklehrkraft in fast 68 % der Fälle „Erklären“ in unterschiedlichen Kontexten, wie beispielsweise „Geduld beim Erklären“ oder „mehrmals erklären“ (S. 3).

Mehrere Studien aus dem Bereich der Unterrichtsforschung erheben die Kompetenz von Lehrkräften auf der Basis von Beurteilungen durch Schüler*innen oder weiterer Personengruppen wie Fachdidaktiker*innen. Beispielsweise beurteilten Schüler*innen in der DESI-Studie (DESI-Konsortium, 2008) die Kompetenzen von Deutschlehrkräften. Während unter anderem gefragt wurde, ob die Lehrkraft für Unterrichtsthemen begeistern oder im Unterricht klare und verständliche Aufgaben stellen könne (DESI-Konsortium, 2008), finden sich jedoch keine Items, die explizit die Erklärkompetenz der Lehrkräfte thematisieren. Allerdings besteht die Vermutung, dass beispielsweise die Fragen zum Aspekt Unterstützung durch die Lehrperson auch Erklärkompetenz berücksichtigen. Auch die TIMS-Studie untersuchte in verschiedenen Ländern unter anderem das Unterrichtserleben aus Sicht von Schüler*innen (Reusser & Pauli, 2003). Diese wurden sowohl zu ihren Einschätzungen über verschiedene Aspekte des Unterrichts, wie Strukturiertheit und Zielklarheit, als auch über wichtige Kompetenzen der Lehrkraft beziehungsweise deren

Erklärkompetenz befragt (Beispielitem: „Unsere Mathematiklehrerin kann gut erklären“, Reusser & Pauli, 2003, S. 89). In einem anderen Vorhaben wurden Zusammenhänge zwischen Beurteilungen von Schüler*innen über die Erklärkompetenz von Lehrkräften und ihren Einschätzungen zur eigenen Aufmerksamkeitsspanne untersucht, um dadurch Leistungserfolge von Schüler*innen zu erklären (Belgard et al., 1968). Gindele und Voss konnten anhand der Daten aus der COACTIV-R-Studie³⁶ (Kunter et al., 2011) einen positiven Zusammenhang zwischen Einschätzungen von Schüler*innen zur Erklärbarkeit ihrer Lehrkräfte und deren pädagogisch-psychologischem Wissen feststellen, wobei die Autorinnen unter Fertigkeiten das Geben guter Erklärungen verstehen (Gindele & Voss, 2017).

Auf einer anderen Ebene, jedoch auch im Kontext von Erklären in Vermittlungskontexten, untersuchte Odora (2014) Einschätzungen von High School Schüler*innen aus Südafrika mit dem Ziel, Hinweise über Kriterien guten Erklärens von Lehrkräften im Fach Technologie zu erhalten. Beurteilt wurden dabei verschiedene Aspekte wie die verwendeten Beispiele während oder kurze Zusammenfassungen am Ende der Erklärung sowie Merkmale der Lehrkraft, zum Beispiel ihre sprecherischen Fähigkeiten (Odora, 2014). Die Expertise von Fachdidaktiker*innen diente beispielsweise dazu, auf Basis von deren Einschätzungen eine Heuristik über Erklärkompetenzen von Lehrkräften im wirtschaftspädagogischen Unterricht zu entwickeln (Schopf & Zwischenbrugger, 2015).

In der Musikpädagogik fehlen Beurteilungen über die Erklärkompetenz von Musiklehrkräften gänzlich. Mit weitreichender Einschränkung ist beispielsweise die Studie von Schulten (1980) über die Wünsche von Schüler*innen zu Methoden des Musikunterrichts zu nennen. Dort finden sich zwei Items zum Erklären: „Der Lehrer benutzt Lieder zum Erklären von Notenwerten, Intervallen usw.“ und „Der Lehrer benutzt Tafel oder Projektor zum Erklären“ (Schulten, 1980, S. 100). Allerdings lassen sich aus den präsentierten Daten aufgrund teils fehlender Werte für das zweite Item oder einer fehlenden Aufschlüsselung der befragten Schüler*innen aus dem Gymnasium keine aussagekräftigen Schlüsse ziehen. Eine umfassende Auseinandersetzung mit der Erklärkompetenz von Musiklehrkräften als Möglichkeit, auf die Qualität von Musikunterricht einzuwirken, könnte mithilfe von Einschätzungen aus unterschiedlichen Perspektiven von am Musikunterricht direkt wie indirekt Beteiligten gelingen. Dies zeigen andere fachfremde Ansätze, in denen unterschiedliche Personengruppen die Erklärkompetenz von Lehrenden beurteilen (z. B. Odora, 2014; Reusser & Pauli, 2003) oder Fachdidaktiker*innen ihre Einschätzung über Erklärkompetenzen einer Lehrkraft im Kontext wirtschaftspädagogischen Unterrichts abgeben (Schopf & Zwischenbrugger, 2015).

2.3 Erklären im Unterricht

Beim Blick in die Forschungsliteratur zum unterrichtlichen Erklären einer Lehrkraft zeigt sich ein ambivalentes Bild. Auf der einen Seite wird innerhalb der Bildungsforschung an

36 Der Buchstabe *R* steht im Akronym für *Referendar*innen* und meint eine Erweiterungsstudie von COACTIV (Kunter et al., 2011).

verschiedenen Stellen davon gesprochen, dass Erklären zu den Kerntätigkeiten einer Lehrkraft im Unterricht gehört (Findeisen, 2017; Fraefel & Scheidig, 2018; Leinhardt, 2001; Pauli, 2015). Auf der anderen Seite beschreiben Schilcher et al. (i. Vorb.) den Forschungsstand zum unterrichtlichen Erklären von Lehrkräften über die unterschiedlichen Fachdidaktiken hinweg als heterogen und trotz einzelner theoretischer wie empirischer Beiträge nach wie vor als defizitär. In der musikpädagogischen Fachliteratur stellt das Erklären im Kontext von Musikunterricht bisher ein randständiges Thema dar. Aufgrund dieses Desiderats richtet sich der Fokus im Folgenden zunächst auf den Stand der Forschung zum Erklären in den Erziehungswissenschaften und anderen Fachdidaktiken, zu der thematisch anschlussfähige musikpädagogische Forschungsliteratur in Beziehung gesetzt wird. Um Erklären im Kontext von Unterricht zu fassen und die terminologische Rahmung (siehe Kap. 2.1) mit inhaltlichen Aspekten auszugestalten, liegt das Interesse zunächst auf Charakteristika, die in eine Definition unterrichtlichen Erklärens münden. Auf Basis dessen werden in einem zweiten Schritt Merkmale guten unterrichtlichen Erklärens einer Lehrkraft entfaltet.

2.3.1 Charakteristika

Zur Konturierung des unterrichtlichen Erklärens dienen zunächst außerunterrichtliche Bereiche, in denen Erklären eine Rolle spielt. Neben Kontexten, wie Hochschullehre oder erlebnispädagogischen Vermittlungsformaten für den Besuch von Museen (z. B. Kim & Crowley, 2010; Knutson & Crowley, 2010), kommt Erklären beispielsweise auch in behördlicher Beratung zum Einsatz (Hohenstein, 2009). Der Kommunikationswissenschaftler Hargie (2013) untersucht Erklären auch in Kontexten wie Medizin, Technologie oder Wirtschaft und Recht. Dabei greift der Autor auf einen weit gefassten Begriff von Erklären zurück, der anhand vielfältiger Ziele deutlich wird: Erklären umfasse nicht nur das Zugänglichmachen von sonst nicht verfügbaren Informationen oder die Demonstration von Techniken, sondern schließe unter anderem auch das Ausdrücken eigener „Meinungen zu bestimmten Einstellungen, Fakten oder Wertvorstellungen“ (Hargie, 2013, S. 265) ein.

Diese weite Definition von Erklären steht jedoch dem Anliegen aus der Linguistik entgegen, Erklären nicht als übergeordnetes Synonym für andere Handlungen wie beispielsweise *Begründen* oder *Instruieren* zu verwenden (Hohenstein, 2009; Klein, 2009; siehe ausführlich Kap. 2.1). Auch bleibt bei den genannten Zielen unklar, ob die Demonstration von Techniken im Sinne Hohensteins (2009) das Ziel verfolgt, Wissen aufzubauen oder zum Handlungsvollzug zu befähigen. Aus den genannten Gründen erweist sich die breite Konzeptualisierung von Erklären nach Hargie (2013) für die vorliegende Arbeit als unpassend. Allerdings unterscheidet der Autor verschiedene Arten des Erklärens und zählt zu den drei wichtigsten (1) das rein verbale Erklären, (2) das Erklären unter Ergänzung von Veranschaulichungen und (3) das *Explaining by Doing*, also das Erklären mit Hilfe der Demonstration von Handlung (Hargie, 2013, S. 264). Alle drei Arten wiederum können im (Musik-)Unterricht vorkommen und unterrichtliches Erklären als

Handeln einer Lehrkraft in mehrere Facetten ausdifferenzieren, auch wenn sie in Hargies (2013) Darstellung bewusst über den unterrichtlichen Kontext hinaus gehen.

Um ein detaillierteres Bild unterrichtlichen Erklärens zu erhalten, sind jedoch weitere Charakteristika notwendig. Offensichtlich ist zunächst, dass unterrichtliches Erklären einer Lehrkraft im Unterricht beziehungsweise innerhalb der Institution Schule stattfindet. Demnach unterscheidet es sich beispielsweise von alltagstypischen Kontexten, in denen etwas erklärt wird, anhand der Voraussetzungen, die aufgrund des institutionellen Rahmens gegeben sind: Im Gegensatz zu einer freiwilligen Situation, bei der eine Person einer anderen beispielsweise den Weg erklärt, vermittelt eine Lehrkraft fachspezifisches Wissen an Schüler*innen, deren Teilnahme am Unterricht aufgrund der Schulpflicht geregelt ist (Vogt, 2009). Weitere Vorgaben, die den schulischen Kontext beeinflussen, liegen auch auf curricularer Ebene. Eine Lehrkraft wählt Inhalte, die im Lehrplan vorgegeben sind, aus und macht sie allen Schüler*innen gleichermaßen zugänglich – das ist gerade in leistungsheterogenen Klassengemeinschaften relevant. Unterrichtliches Erklären erfolgt also immer unter bestimmten Rahmenbedingungen, die gesetzt sind.

Beim Erklären in Lehr-Lernkontexten – Unterricht ist als eine Form davon zu betrachten – wird auch von *instruktionalen Erklärungen* gesprochen, die zu den häufigsten Elementen instruktionaler Lehr-Lern-Settings gezählt werden (Renkl et al., 2006). Renkl (2017) definiert sie als „Erklärungen, die in Lehr-Lern-Kontexten gegeben werden, um das Wissen und Verstehen der Lernenden zu erweitern“ und präzisiert hinsichtlich des Umfangs, dass es sich meist um „kürzere Darlegungen einer Lehrkraft“ (S. 799) handle. Die Begriffe *unterrichtlich* und *instruktional* können aus mehreren Gründen synonym verwendet werden. Sie sind beide nicht nur in den Bildungskontext eingebettet, sondern auf der gleichen Ebene angesiedelt: Instruktionales Erklären ist ein Element instruktionaler Lehr-Lernsettings (Renkl et al., 2006; Vogt, 2009) und unterrichtliches Erklären ein Element von Unterricht – beide können auf einer Mikroebene verortet werden und stehen im Verhältnis zu Lehr-Lernsituationen oder Unterricht auf einer Makroebene. Neben der meist kürzeren Dauer einer instruktionalen Erklärung können weitere Prinzipien, wie die Berücksichtigung des Vorwissens der Adressat*innen oder die Fokussierung der erklärenden Person auf den zu erklärenden Gegenstand (Leinhardt, 2001), für unterrichtliches Erklären fruchtbar gemacht werden (siehe ausführlich Kap. 2.3.2). Sie sind von Interesse, da verschiedene Untersuchungen bereits die Wirksamkeit von Prinzipien instruktionalen Erklärens nachweisen konnten (Nückles et al., 2005; Stein & Kucan, 2010).

Leinhardt (2001) unterscheidet *instructional explanations* von *common explanations* (allgemeine Erklärungen; meist Antworten auf direkte und einfache Fragen), *self-explanations* (Selbsterklärungen) und *disciplinary explanations* (eine Disziplin betreffende Erklärung; unterschiedliche Gebiete wie Psychologie, Geschichte oder Mathematik stellen spezifische Fragen und suchen nach unterschiedlichen Antworten). Instruktionale Erklärungen sollen dabei einerseits die Lücke zwischen allgemeinen Erklärungen und disziplinspezifischen Erklärungen schließen und unterscheiden sich andererseits von den drei anderen Erklärarten dadurch, dass sie konkret in Lehr-Lernsituationen vorkommen und damit den Bildungs- beziehungsweise Unterrichtskontext fokussieren (Leinhardt, 2001). Gerade im Vergleich zwischen unterrichtlichen und alltäglichen Erklärungen sind diese

aus linguistisch-didaktischer Perspektive auch anhand der Komplexität des zu erklärenden Gegenstandes zu unterscheiden (Kotthoff, 2009). Diese Komplexität des Gegenstandes nimmt Einfluss auf die Handlung des Erklärens, die ebenfalls als komplex beschrieben wird (Findeisen, 2017; Wörn, 2014). Spreckels (2009b) vermutet, dass diese Komplexität der Handlung zum Forschungsdesiderat unterrichtlichen Erklärens beiträgt.

Ein weiteres Charakteristikum unterrichtlichen Erklärens, das jedoch auch in außerunterrichtlichen Kontexten wie behördlicher Beratung oder Gesprächen zwischen Ärzt*innen und Patient*innen eine Bedeutung hat (Hargie, 2013; Hohenstein, 2009), ist die vorherrschende Wissensasymmetrie zwischen der Lehrkraft als erklärender Person und den Schüler*innen als Adressat*innen der Erklärung (Findeisen, 2017; Kiel, 1999; Leinhardt, 2001; Vogt, 2009). Auf Seiten der Lehrkräfte setzt dies ausreichend fachliches Wissen voraus, um bestimmte Sachverhalte zu erklären (Findeisen, 2017). In diesem Zusammenhang unterscheiden Krauss et al. (2017) hinsichtlich des fachlichen Wissens von Lehrkräften zwischen Alltagswissen, Schulwissen, vertieftem Hintergrundwissen zum Schulstoff und universitärem Wissen, wobei das Fachwissen auf Ebene zwei und drei operationalisiert wird. Fehlendes Wissen oder auch Verständnisfehler auf Seiten der Schüler*innen hingegen evozieren überhaupt erst den Bedarf einer Erklärung (Wagner & Wörn, 2011). Allerdings merkt Hohenstein (2009) an, dass die Adressat*innen einer Erklärung trotzdem über gewisse Voraussetzungen verfügen müssen, damit die erklärende Person Wissen und Verständnis auf- beziehungsweise ausbauen kann.

Dass Schüler*innen etwas verstehen, wird nicht nur als allgemeines Ziel von Unterricht (z. B. Diederich & Tenorth, 1997), sondern auch als grundlegende Funktion unterrichtlichen Erklärens bezeichnet (Findeisen, 2017; Kiel, 1999; Neumeister, 2011; Shulman, 1986) und kann daher ebenfalls als Charakteristikum angesehen werden. Dieses vor allem kognitive Ziel wird jedoch auch breiter gefasst, indem Erklären – ganz im Sinne des Kompetenzerwerbs – über den Wissensaufbau hinaus auch auf das Handeln zielt (Kiel, 1999; Klein, 2009). Zwar werden Erklären und Verstehen als Komplement angesehen (Bartelborth, 2007; Kiel, 1999), jedoch sind beide Prozesse grundsätzlich unabhängig voneinander zu betrachten, da nach Klein (2009) die perlokutive Wirkung des illokutionären Akts Erklären nicht gegeben sein muss: So könne man auch von *Erklären* sprechen, wenn jemand etwas erklärt, obwohl auf Adressat*innenseite kein Verstehen festgestellt werden kann (Illokution); im Gegensatz dazu könne beispielsweise jedoch nur von *Überzeugen* geredet werden, wenn eine andere Person auch tatsächlich überzeugt wurde (Perlokution).

In der bildungswissenschaftlichen Fachliteratur wird kontrovers diskutiert, ob Interaktion ein Charakteristikum unterrichtlichen Erklärens darstellt. Leinhardt (2001) betont das kommunikative Element von Erklären zwischen Lehrkraft und Schüler*innen: „Instructional explanations are designed to explicitly teach – to specifically communicate some portion of the subject matter to others, the learners“ (S. 340). Im Gegensatz dazu fasst Wörn (2014) Erklären nicht als Interaktion, sondern als Handlung auf. Die Autorin begründet dies damit, dass Interaktion laut Duden „die Beteiligung mindestens zweier Personen“ impliziere, das Sich-Selbst-Erklären dieser Logik folgend „keine Erklärungs-handlung“ wäre und dies wiederum im Kontext von Unterricht „als nicht begründbare

Einschränkung verstanden“ werde (Wörn, 2014, S. 20). In diesem Zusammenhang zeigen verschiedene Untersuchungen, dass die Förderung von Selbsterklärungen der Lernenden unter bestimmten Rahmenbedingungen mitunter effektiver sein kann als das Geben instruktionaler Erklärungen (Schworm & Renkl, 2006; Stark et al., 2001). Betrachtet man unterrichtliches Erklären jedoch ausschließlich aus der Perspektive einer Lehrkraft, wie es in der vorliegenden Arbeit bereits eingegrenzt ist (*unterrichtliches Erklären einer Lehrkraft*, siehe Kap. 2.1), stellt Interaktion ein entscheidendes Charakteristikum dar (Findeisen, 2017; Kiel 1999; Prediger & Erath, 2014). Nachdem Selbsterklärungen von Schüler*innen weder Thema dieser Untersuchung sind, noch das Charakteristikum der Wissensasymmetrie zwischen erklärender Person und Adressat*innen bei Selbsterklärungen passend erscheint, erweist sich die Definition von Wörn (2014) für die vorliegende Arbeit als ungeeignet.

Im Kontext von Interaktion wird an anderer Stelle auch von Erklären als interaktivem Kommunikationsprozess gesprochen (Klein, 2009). Dies führt zur Frage, wie sich ein Erklärprozess gestaltet. Vogt (2009) beschreibt für den Literaturunterricht vier Phasen, wobei sich der Prozess über zwei Unterrichtsstunden erstreckt. Mit Verweis auf das *Grundmodell des Erklärens* nach Kiel (1999) unterteilt der Autor den Prozess in Aufmerksamkeitsausrichtung, Phänomenisolierung, abstrahierende Fokussierung und Exemplifizierung (Vogt, 2009). Diese Einteilung kann für die vorliegende Arbeit aus verschiedenen Gründen nicht verwendet werden. Erstens veranschlagt der Autor für die Phase der Aufmerksamkeitsausrichtung eine ganze Schulstunde (Vogt, 2009), was der eher kurzen Dauer aus der Definition instruktionalen Erklärens nach Renkl (2017) entgegensteht. Zweitens schränkt Vogt (2009) selbst ein, dass Teile der zweiten Phase der Phänomenisolierung „nicht unbedingt dem Erklärprozess zugeordnet werden“ (S. 205) können. Drittens beschreibt er die letzte Phase als Möglichkeit für die Schüler*innen, das erworbene Wissen anzuwenden (Vogt, 2009).

Diese drei Aspekte führen zur Frage, welcher Begriff von Erklären zugrunde liegt, beziehungsweise wie sich *Erklären von Unterrichten* unterscheidet. Es ist zu vermuten, dass diese Prozesseinteilung auf einem Begriff von Erklären beruht, der nicht das Handeln einer Lehrkraft fokussiert, sondern ein grundsätzliches Ziel von Unterricht beschreibt: Inhalte zu vermitteln. Eine andere Form eines Erklärprozesses legen Wagner und Wörn (2011) für den Mathematikunterricht vor, der auch an anderer Stelle aufgegriffen und ausführlich beschrieben wird (Findeisen, 2017). Bei diesem gehen dem Erklären sowohl ein Erklär Anlass, also ein Grund beziehungsweise eine Notwendigkeit für eine Erklärung, als auch eine Erklärinitiierung, die Signalisierung des Bedürfnisses nach Erklärung, voraus (Wagner & Wörn, 2011). Die Erklärcoda schließt den Prozess ab, indem sich die Lehrkraft beispielsweise rückversichert, ob die Schüler*innen alles verstanden haben (Wagner & Wörn, 2011). Die Erklärungen, die in der vorliegenden Arbeit verhandelt werden, folgen in ihrer Grundstruktur eher dieser Prozessdarstellung.

Anhand des Erklär Anlasses lassen sich zwei weitere Charakteristika unterrichtlichen Erklärens verdeutlichen, die im Vollzug unterschiedliche Formen annehmen. Erklären kann sowohl spontan und unvorbereitet als auch geplant und vorbereitet erfolgen, wobei sich beide Varianten in mehreren Aspekten wie zeitlicher Dauer, Adressatenbezug oder

auch dem Anlass unterscheiden können (Wagner & Wörn, 2011). Während unvorbereitete Erklärungen auf Fehler oder Nachfragen der Schüler*innen reagieren, eher kurz und prägnant sind und sich teilweise auch nur an einzelne Lernenden richten können, dienen vorbereitete Erklärungen der Einführung neuer Inhalte, sind oft etwas ausführlicher und richten sich grundsätzlich an alle Schüler*innen (Wagner & Wörn, 2011). Nach Wagner und Wörn (2011) bieten geplante Erklärungen den Vorteil, Materialien gezielt auszuwählen und die Erklärung zu durchdenken, wohingegen eine spontane Erklärung im Unterrichtsprozess meist als Reaktion auf eine Frage oder einen Fehler einer Schülerin oder eines Schülers erfolgt und in der Regel ohne unterstützende Materialien und ohne Zeit für eine umfassende Vorbereitung und Planung des Erklärens auskommen muss (Wagner & Wörn, 2011).

In einem qualitativen Setting untersuchte Spreckels (2009b) spontane Erklärungen im Unterricht zweier Lehramtsstudierender und einer erfahrenen Lehrkraft. Von Interesse war, inwieweit die Lehrpersonen in verschiedenen Unterrichtssequenzen jeweils mit Rückmeldungen der Schüler*innen umgehen und spontan handeln, beziehungsweise in welcher Qualität sie *ad hoc* erklären konnte. Die Autorin konnte beobachten, dass die erfahrene Lehrkraft bei Zwischenfragen viel ausführlicher auf die Bedürfnisse der Schüler*innen einging und mehr erklärte, während sich die beiden angehenden Lehrkräfte stärker am Lernziel der Stunde orientierten und nur äußerst knapp auf die Rückfragen eingingen, beziehungsweise nicht erklärt haben (Spreckels, 2009b). Bewusst geplante und vorbereitete Erklärungen bieten nach Wagner und Wörn (2011) den Vorteil, „im Vorfeld des Unterrichts in didaktischer sowie methodischer Hinsicht vorbereitet“ zu werden, wobei auch „noch genügend Flexibilität für Erkläralternativen und das Einschleusen von Zwischenschritten vorhanden sein“ (S. 22) sollte.

Eine weitere Unterscheidung, nach der unterrichtliches Erklären charakterisiert werden kann und in fachdidaktischer Forschung untersucht wird (z. B. Prediger & Erath, 2014; Schwab, 2009; Seedhouse, 2009; Wörn, 2014), stellt die Form der Kommunikation zwischen Lehrkraft und Schüler*innen dar. Diese kann sich nicht nur mündlich, sondern auch schriftlich in Form von Schulbuchtexten oder Bedienungsanleitungen vollziehen (Leinhardt, 2001; Wittwer & Renkl, 2008). Der Vorteil schriftlicher Erklärungen liegt beispielsweise darin, dass die Adressat*innen Texte mehrmals lesen, schwierige Stellen intensiver bearbeiten und generell in ihrem individuellen Tempo fortschreiten können (Wellenreuther, 2013). In unterrichtlichen Situationen erklärt eine Lehrkraft jedoch zu meist mündlich (Neumeister, 2011; Spreckels, 2009a). Dadurch hat sie die Möglichkeit, das Erklären in Verbindung mit Visualisierungen, Grafiken oder Analogien flexibel anpassen zu können (Wellenreuther, 2013). Zudem wird angenommen, dass gerade das spontane Erklären in den allermeisten Fällen mündlich und als Reaktion auf Nachfragen oder Äußerungen von Schüler*innen erfolgt. Es ist jedoch zu berücksichtigen, dass gerade lange mündliche Erklärungen einer Lehrkraft von Schüler*innen ausdauernde Aufmerksamkeit erfordern (Wellenreuther, 2013). Eine Variante, die Vorzüge beider Erklärvarianten kombinieren kann, stellen Erklärvideos dar. Eine Erklärung kann dabei nicht nur mündlich erfolgen und durch Visualisierungen ergänzt, sondern auch mehrmals

angesehen werden; jedoch kann die Lehrkraft beziehungsweise die erklärende Person auf der anderen Seite bei Verständnisproblemen nicht befragt werden (Frei et al., 2020).

Auf Grundlage der genannten Charakteristika wurde eine Definition unterrichtlichen Erklärens einer Lehrkraft auf Basis von Literaturstudien entwickelt. Sie liegt dem Projekt FALKE-q (Schilcher et al., 2021) zugrunde und stellt demnach auch den Ausgangspunkt für das vorliegende Forschungsvorhaben (FALKE-q-Mu; Frei et al., i. Vorb.) dar:

Unterrichtliches Erklären einer Lehrkraft ist ein komplexer, vorbereiteter oder sich situativ ergebender sowie interaktiver Kommunikationsprozess, der auf eine Fähig- bzw. Fertigkeitentwicklung sowie die Initiierung eines Verstehensprozesses bei Adressatinnen und Adressaten abzielt. (Lindl et al., 2019, S. 131; siehe dazu Kiel, 1999; Leinhardt, 2001; Prediger & Erath, 2014; Wagner & Wörn, 2011).

Die Definition hat den Anspruch fachübergreifender Gültigkeit und ist in Anlehnung an Neumeister (2011) dem Erklären₂ zuzuordnen (siehe Kap 2.1). Adressiert werden komplexe Prozesse, die in den Alltagssituationen von Erklären₁ nicht zwingend abgebildet werden (siehe Neumeister, 2011). Ähnlich wie dem Angebot-Nutzungs-Modell zur Unterrichtsqualität können dadurch jedoch keine fachspezifischen Feinheiten abgebildet werden. Um auf einer grobkörnigen Ebene Unterschiede zwischen mehreren Unterrichtsfächern beziehungsweise Fachgruppen aufzuzeigen, bietet sich eine Typisierung von Erklären an, die mehreren deutschdidaktischen Untersuchungen zugrunde liegt (z. B. Gaier, 2020; Neumeister, 2011). Die Klassifikation geht auf Brown und Hatton (1982) zurück und bietet durch die Unterscheidung zwischen Erklären-Was, Erklären-Wie und Erklären-Warum „eine in linguistischen Kategorien fundierte kleine Typologie“ (Klein, 2009, S. 26).

Erklären-Was komme dabei prototypisch bei Wort- oder Begriffserklärungen zum Einsatz und habe das Ziel, „(Bestimmungs-)Merkmale für die jeweilige Kategorie zu finden, ausgelöst durch die Frage *Was ist ein(e)...*“ (Neumeister, 2011, S. 45). Erklären-Wie hingegen beschreibt Handlungen, Tätigkeiten und Abläufe, wie zum Beispiel musikinstrumentale Spieltechniken (Klein, 2009). Allerdings sind bezüglich der Unterscheidung von Handlungserklärungen und Instruktionen, die per Definition zum Handlungsvollzug befähigen sollen (Hohenstein, 2009), Unklarheiten festzustellen. Nach Rathausky (2010) werden *Instruieren* und *Erklären(-Wie)* in der sprachwissenschaftlichen Forschungsliteratur sowohl synonym zueinander als auch unabhängig voneinander verwendet. In Bezug auf das vorliegende Forschungsvorhaben wird noch ausführlich erörtert, welche Unterschiede zwischen dem Erklären-Wie und dem Instruieren vorliegen und wie sich diese im Kontext von Musikunterricht ausgestalten können (siehe Kap. 2.4.2). Der dritte Typ, Erklären-Warum, expliziert das „Zustandekommen eines Sachverhalts“ (Klein, 2009, S. 30). Die dafür notwendigen Bedingungen sowie die logisch-kausale Ableitung eines zu erklärenden Phänomens lassen Zusammenhänge zum wissenschaftstheoretischen Begriff von Erklären erkennen (Klein, 2009), wie er beispielsweise dem Schema von Hempel und Oppenheim (1948) zugrunde liegt. Allerdings stellt Klein (2009) typologische und sprachstrukturelle Unterschiede fest, weshalb Erklären-Warum

demnach nicht nur eine Betrachtung kausaler Ableitungen bedeutet, sondern Erklärungen meint, die auf eine Warum-Frage folgen.

Inwieweit die Typologie in Erklären-Was, Erklären-Wie und Erklären-Warum in Bezug auf Charakteristika unterrichtlichen Erklärens geeignet ist, um unterschiedliche Schwerpunkte zwischen verschiedenen Fächern beziehungsweise Fachgruppen zu skizzieren, muss zumindest bezweifelt werden. Zwar scheinen grundsätzliche Tendenzen vorzuliegen, beispielsweise wird Erklären-Warum häufig mit einem (wissenschaftstheoretischen) kausal-deduktiven Erklärschema in Verbindung gebracht, das vor allem für mathematisch-naturwissenschaftliche Fächer charakteristisch ist. Allerdings zeigte Gaier (2020), dass Erklären-Warum durchaus in geisteswissenschaftlich geprägten Fächern wie Deutsch eine Rolle spielt. Auch im Musikunterricht ist die Erklärung denkbar, *warum* beispielsweise der Leitton in der Regel – dies ist eine entscheidende Anmerkung – in den Grundton aufzulösen ist. Diese *Regel* stellt jedoch kein Naturgesetz dar, sondern etablierte sich als eine über Jahrhunderte gewachsene Entwicklung westlicher Musikkultur. Ein Erklären-Warum ist daher als linguistische Typologie für den Musikunterricht dann denkbar, wenn es nicht nur durch einen kausal-deduktiven Erklärungs begriff definiert ist.

Ein Erklären-Wie bringt Klein (2009) vor allem mit sportlich-musischen Fächern in Verbindung, in deren Vermittlungskontexten vielfach Handlungsanweisungen gegeben werden und ein rein kognitiver Nachvollzug auf Seiten der Adressat*innen für den eigenen Handlungsvollzug nicht ausreicht. Demnach könnte eine Musiklehrkraft sowohl erklären, *wie* ein Instrument gespielt wird, *wie* dabei Klänge entstehen oder *wie* eine Dur-Tonleiter aufgebaut ist – die Unterscheidung zum Instruieren liegt diesbezüglich darin, dass eher Wissensaufbau statt Handlungsvollzug intendiert ist. Erklären-Wie dabei vor allem sportlich-musischen Fächern zuzuordnen, greift ebenfalls zu kurz. Im Fach Mathematik beispielsweise kann anhand verschiedener Rechenschritte erklärt werden, *wie* der jeweilige Lösungsweg abläuft. Beim Erklären-Was wiederum sind im Kontext der Einführung beziehungsweise Erklärung von Begriffen und dahinterliegenden Konzepten zwar unterschiedliche Ausprägungen, jedoch ein ähnlich verteiltes Auftreten zwischen Schulfächern zu erwarten (Klein, 2009). Sie kommen im Musikunterricht beispielsweise vor, wenn erklärt wird, *was* der Begriff Terz im Kontext von musiktheoretischen Inhalten bedeuten kann oder *was* eine Kadenz ist.

In der musikpädagogischen Literatur finden sich nahezu keine expliziten Hinweise zu Charakteristika unterrichtlichen Erklärens einer Musiklehrkraft. Nur an vereinzelten Stellen scheint durch, dass eine Musiklehrkraft beispielsweise nicht erklären solle, wenn gemeinsam musiziert wird (Eberhard, 2010). Im Kontext des Instrumentalunterrichts kommt Erklären vor allem im Rahmen der Darstellenden Methode³⁷ (Ernst, 1991) zum Einsatz, wobei es vor allem bei Verständnisproblemen und zumeist erst „auf eine Nachfrage der Schüler*innen hin“ geschehe (Dartsch, 2014, S. 169; siehe auch Fuchs, 2015). Es wird angenommen, dass solche Verständnisprobleme bereits durch eine Lehrkraft mit

37 Ernst (1991) nennt sechs methodische Verfahren für den Instrumentalunterricht. Dazu zählen (1) das Erarbeitende Verfahren, (2) die Modell-Methode, (3) das Darstellende Verfahren, (4) das Aufgebende Verfahren, (5) das Entdeckenlassende Verfahren und (6) die Dialog-Methode (siehe für eine Zusammenfassung Dartsch, 2014, S. 169–173).

einer kurzen Erklärung vorentlastet werden können – also bereits von der Lehrkraft ausgehend und nicht erst auf Nachfrage von Schüler*innen vollzogen wird. Auch alle weiteren, bisher erläuterten Charakteristika, wie das zugrundeliegende Wissensdefizit oder die spontane beziehungsweise vorbereitete Umsetzung sollten für den Musikunterricht gültig sein. Eine ausführliche Aufarbeitung der Fragen, in welchen Kontexten und in welcher Weise im Musikunterricht erklärt wird, ist Inhalt des Kapitels 2.4.

2.3.2 Qualitätsmerkmale

Hinsichtlich des Diskurses über Qualitätsmerkmale unterrichtlichen Erklärens einer Lehrkraft zeigt sich beim Blick in die verschiedenen Fachdidaktiken ein heterogenes Bild (Schilcher et al., i. Vorb.). Die Musikpädagogik zählt dabei zu den Disziplinen, in denen bisher noch keine empirische Forschungsarbeiten und erst vereinzelt theoretische Beiträge vorliegen (z. B. Frei et al., 2022). Daher wird auch an dieser Stelle vor allem der fachübergreifende Stand der Forschung zu Merkmalen unterrichtlichen Erklärens dargestellt und mit dem musikpädagogischen Diskurs in Beziehung gesetzt. In Kapitel 2.4 folgt dann eine musikspezifische Einordnung. Analog zur synonymen Verwendung der zwei Termini *Unterrichtsqualität* und *guter Unterricht* werden für die folgende Darstellung auch *Erklärqualität* und *gutes Erklären* bedeutungsgleich verwendet. Dadurch ist es auch auf sprachlicher Ebene leichter möglich, die Diskussion um Qualitätsmerkmale unterrichtlichen Erklärens mit dem Diskurs um Unterrichtsqualität zu verbinden.

Der folgende Abschnitt entfaltet, welche Merkmale für gutes unterrichtliches Erklären in der Fachliteratur benannt und wie diese konzeptualisiert werden. Dabei wird herausgearbeitet, ob diese Kriterien in der verwendeten Fachliteratur normativ gesetzt oder bereits empirisch untersucht sind. Hinsichtlich der empirischen Validität ist weiterhin zu unterscheiden: Sind auf Basis der Studien Aussagen über die Wirksamkeit bestimmter Aspekte guten Erklärens möglich – dies ist für das Merkmal Adressatenorientierung möglich (siehe z. B. Nückles et al., 2005; Stein & Kucan, 2010; Wittwer & Renkl, 2008). Oder werden Visualisierungen beispielsweise von Schüler*innen als wichtiges Element beurteilt, um das Erklärte zu verstehen (Kulgemeyer & Tomczyszyn, 2015), und können auf Basis einer solchen empirischen Erkenntnis als Merkmal guten Erklärens angesehen werden.

Der Blick in die didaktische und erziehungswissenschaftliche Fachliteratur bringt eine Fülle an Merkmalen guten Erklärens im Unterricht hervor, die zum Teil normativ gesetzt, in manchen Fällen aber auch empirisch validiert sind. Beide Kategorien dienen der folgenden Darstellung als Orientierung, wobei nur eine Auswahl von Merkmalen vorgestellt wird, die in Bezug auf unterrichtliches Erklären immer wieder genannt werden (für eine ausführliche Literaturübersicht siehe Findeisen, 2017, S. 50–52). Auf Basis empirischer Erkenntnisse sollen beim Erklären

- das Vorwissen der Adressat*innen berücksichtigt,
- angemessene Visualisierungen und passende Beispiele verwendet,
- der zu erklärende Gegenstand fachlich korrekt präsentiert,

- in Form von Bildern oder Metaphern repräsentiert,
- sich auf das Wesentliche beschränkt,
- die wichtigsten Inhalte am Ende zusammengefasst,
- abwechslungsreich vorgetragen und
- eine angemessene Sprache benutzt werden (z. B. Findeisen, 2017; Kulgemeyer & Schecker, 2013; Leinhardt, 2001; Schopf & Zwischenbrugger, 2015; Wittwer & Renkl, 2008).

Darüber hinaus sollen beim Erklären auf Basis normativer Qualitätsvorstellungen

- ein Bezug zur Lebenswelt der Schüler*innen hergestellt,
- der zu erklärende Gegenstand sinnvoll didaktisch reduziert,
- seine Kernprinzipien herausgearbeitet und
- dabei generell schrittweise vorgegangen werden (z. B. Kiel, 1999; Lehner, 2018; Scheffel, 2019).

Die folgende Darstellung gliedert die genannten Merkmale inhaltlich nach den Kategorien Adressatenorientierung, Strukturiertheit, Sprachliche Verständlichkeit, Sprech- und Körperausdruck sowie Repräsentationen (siehe Frei et al., 2022). Die Auswahl entspricht dabei einer normativen Setzung und ist zum Großteil – außer dem Einsatz von Repräsentationen – das Ergebnis eines auf Literaturrecherche basierenden fachübergreifenden Diskurses zwischen insgesamt 14 Disziplinen, die im Rahmen des Projekts FALKE-q zusammengearbeitet haben (Schilcher et al., 2021).³⁸ Die Überschneidungen zwischen Merkmalen guten Unterrichts (Helmke, 2017) und Merkmalen guten Erklärens waren dabei kein primäres Ziel der Konzeptualisierung, sind aufgrund der engen Verknüpfung zwischen Erklären und Unterrichten jedoch naheliegend.

38 Dabei wurden Merkmale guten Erklärens identifiziert, denen einerseits fachübergreifende Gültigkeit und andererseits in jedem Fach eine grundlegende Relevanz attestiert wird. Ziel war außerdem, auf Basis der folgenden Merkmale eine fachübergreifende Modellierung von Qualitätsaspekten des Erklärens realisieren zu können: Adressatenorientierung, Strukturiertheit, Sprachliche Verständlichkeit sowie Sprech- und Körperausdruck (Lindl et al., 2019). Auch die Verwendung von Visualisierungen erachteten alle Disziplinen als bedeutsam, wenngleich sie diesen Aspekt aufgrund eines fehlenden einheitlichen Verständnisses über Form und Art der Visualisierung in der fachübergreifenden Konzeptualisierung ausgespart haben (Lindl et al., 2019). Da verschiedene Repräsentationsformen wie Analogien, Beispiele oder Visualisierungen weitestgehend übereinstimmend als Merkmale guten Erklärens gelten (z. B. Findeisen, 2017; Schilcher et al., 2017), werden sie in der folgenden Ausführung sowohl allgemein und fachübergreifend dargestellt als auch mit Blick auf Musikunterricht weiter ausdifferenziert. Da akustische Repräsentationsformen für das Erklären im Musikunterricht und daher für den vorliegenden Beitrag von großem Interesse sind und dazu kein genuin fachübergreifender Diskurs besteht, werden sie erst in Kap. 2.4.2 entfaltet.

Die Orientierung am Wissens- und Kenntnisstand der Schüler*innen als Adressat*innen wird nicht nur von Seiten der Bildungsadministration als grundlegende fachdidaktische Kompetenz angehender Lehrkräfte gefordert (KMK, 2019), sondern auch in der Fachliteratur als ein zentraler Aspekt guten Erklärens im Unterricht festgeschrieben (Brown & Hatton, 1982; Findeisen, 2017; Kiel, 1999; Wagner & Wörn, 2011; Wittwer & Renkl, 2008). Dabei wird deutlich, dass die Adressatenorientierung eher als grundlegendes Prinzip beziehungsweise als Leitgedanke unterrichtlichen Erklärens verstanden wird (Wagner & Wörn, 2011; Wittwer & Renkl, 2008) – ähnlich der grundlegenden Prinzipien Schülerorientierung oder Passung als wichtiges Kriterium für die Qualität von Unterricht (Helmke, 2017). Aus diesem Grund wird Adressatenorientierung auf einer Metaebene als eine Art Leitprinzip unterrichtlichen Erklärens angesiedelt (Wagner & Wörn, 2011). Aspekte, die für ein an den Adressat*innen orientiertes Erklären berücksichtigt werden müssen, sind beispielsweise das Vorwissen und die Lebenswelt der Schüler*innen sowie die Komplexität des zu erklärenden Gegenstandes. Im Handeln der Lehrkraft kann die Orientierung an den Adressat*innen beispielsweise anhand der didaktischen Reduktion, dem hergestellten Bezug zur Lebenswelt der Schüler*innen oder auch durch passende Sprache⁴⁰ realisiert werden.

Nach Wittwer und Renkl (2008) sind insbesondere solche Erklärungen wirksam, die an das Vorwissen der Adressat*innen anknüpfen (siehe z. B. Nückles et al., 2005). Auch in der fachdidaktischen Forschung stellt die Orientierung am Vorwissen der Schüler*innen ein wesentliches Element adressatenorientierten Erklärens dar (z. B. Findeisen, 2017; Kulgemeyer & Tomczyszyn, 2015; Schopf & Zwischenbrugger, 2015). Anhaltspunkte für den Vorwissensstand von Schüler*innen sind für eine Lehrkraft zum einen auf abstrakter Ebene in Form curriculärer Vorgaben über das zu erwartende Wissen in der jeweiligen Jahrgangsstufe zu finden. Zum anderen kann der konkrete Wissens- beziehungsweise Kenntnisstand der Schüler*innen einer Klasse beziehungsweise die Orientierung an einem „fiktiven oder auch realen Durchschnittsschüler“ (Weinert, 1997, S. 51) als Maßstab dienen. Allerdings werden dafür hinreichende diagnostische Kompetenzen von Lehrkräften als notwendig erachtet (Wittwer & Renkl, 2008). Zudem wird die Nähe zu Merkmalen von Unterrichtsqualität auch dadurch deutlich, dass das Konzept der *Passung* beziehungsweise des *Adaptiven Unterrichts* als „Sammelbegriff für unterschiedliche Varianten des Umgangs mit Heterogenität“ (Helmke, 2017, S. 248) fungiert.

Hinsichtlich der Relevanz für unterrichtliches Erklären weist Findeisen (2017) auf ein Dilemma hin, das je nach Grad an Heterogenität der Lerngruppe mehr oder weniger schwer wiegen kann: Die Autorin argumentiert anhand mehrerer Studien, dass Erklärungen, die sich am Vorwissen der lernschwachen Schüler*innen einer Lerngruppe

39 Dieser Begriff wird so im Projekt FALKE-q verwendet. Um eine einheitliche Bezeichnung in allen Forschungsberichten zu gewährleisten, wird diese auch in der vorliegenden Arbeit verwendet. Gemeint sind dabei stets Adressaten und Adressatinnen.

40 Sprachliche Verständlichkeit wird in der vorliegenden Darstellung als eigenes Merkmal guten Erklärens aufgefasst.

orientieren, „ihre Effektivität bei Lernenden mit erweitertem Vorwissen verlieren und ggf. sogar negative Auswirkungen haben“ (Findeisen, 2017, S. 37). Auch Renkl et al. (2006) nehmen an, dass sich die Wirksamkeit einer Erklärung für alle Adressat*innen nur dann entfalten kann, wenn das Angebot der Lehrkraft mit dem Bedarf der Lernenden und damit deren jeweiligen individuellen Wissensständen in optimaler Balance ist. In diesem Zusammenhang verweisen die Autor*innen auf den so genannten *expert blind spot* (siehe z. B. Nathan & Petrosino, 2003), wonach das hohe Fachwissen von Lehrkräften dazu führen kann, dass sie sich „bei der Gestaltung der Lerninhalte nicht an den Bedürfnissen und dem Vorwissen der Lernenden orientieren, sondern sich von der eigenen fachlich geprägten Wahrnehmung des Lerngegenstands leiten lassen“ (Renkl et al., 2006, S. 207).

Mit Blick auf den Musikunterricht und das Unterrichtshandeln von Musiklehrkräften wird dem Umgang mit einem heterogenen Wissens- und Kenntnisstand der Schüler*innen eine große Bedeutung beigemessen (Linn, 2017; Niessen, 2013). Nach Jank (2013) ergeben sich diesbezüglich gerade für den Musikunterricht der Sekundarstufe besondere Anforderungen für Lehrkräfte: Einerseits kann das große Angebot außerschulischer Lernangebote in Form von musikalischer Früherziehung, Instrumental- und Gesangsunterricht sowie der Mitwirkung in Chören oder Orchestern zu einer größeren Heterogenität bei musikalischen Kompetenzen von Schüler*innen führen (Jank, 2013). Diese unterschiedlichen musikalischen Vorerfahrungen sind zentrale Determinanten des Unterrichtsgeschehens im Fach Musik (Linn, 2017). Andererseits tragen auch unterschiedliche institutionellen Bedingungen, wie beispielsweise häufig ausfallender oder fachfremd erteilter Musikunterricht im Primarbereich zu einer verstärkten Heterogenität innerhalb der Schülerschaft bei (Lehmann-Wermser et al., 2020).

Um die Komplexität eines zu erklärenden Gegenstandes auf das Niveau des Vorwissens einer Gruppe von Schüler*innen zu reduzieren, kann das Prinzip der didaktischen Reduktion angewendet werden (Grüner, 1967). Dabei ist es für das unterrichtliche Erklären nicht nur relevant, die Komplexität eines Themas, sondern beispielsweise auch die Stofffülle zu reduzieren (Lehner, 2020). Nach Schopf und Zwischenbrugger (2015) ergibt sich der Schwierigkeitsgrad einer Erklärung – ein angemessener Schwierigkeitsgrad ist auch ein grundsätzliches Merkmal guten Unterrichts (Ditton, 2000) – für jede Schülerin und jeden Schüler aus ihrem eigenen Vorwissen über das erklärte Thema sowie der didaktischen Reduktion von Komplexität und Stofffülle durch die Lehrkraft. Für den Musikunterricht sieht Oebelsberger (2019) das Prinzip der didaktischen Reduktion – die Autorin verwendet vorzugsweise den Begriff Elementarisierung – vor allem als relevant für die Vermittlung musiktheoretischer Inhalte im Unterricht, wenngleich sie in diesem Kontext betont, dass die Musik „nicht durch verfehlte Elementarisierung, Vereinfachung entwertet“ (S. 93) werden darf.

Ein weiterer Parameter von Adressatenorientierung als Merkmal unterrichtlichen Erklärens und gleichzeitig auch didaktisches Unterrichtsprinzip ist der Grad an „Verknüpfungen mit dem Alltag und der Lebenswelt“ (Helmke, 2017, S. 223; siehe auch Wellenreuther, 2013) der Schüler*innen. Schopf und Zwischenbrugger (2015) empfehlen beim Erklären im Unterricht, schülernahe Beispiele zu verwenden, die den Schüler*innen bekannt sind und im besten Fall einen „persönlichen Bezug aus ihrer Lebenswelt“ (S. 31)

aufweisen. In seiner theoretischen Auseinandersetzung mit dem Erklären als didaktisches Handeln fordert Kiel (1999) beim Erklären als Aushandeln von Wissen, dass der zu erklärende Gegenstand einen „deutlichen Bezug zur Lebenswelt und zur Lebenspraxis“ (S. 227) hat. Zu bedenken ist hier jedoch, dass diese Kategorie didaktischen Handelns ein stark schülerorientiertes Vorgehen als Grundprinzip verfolgt, bei dem im besten Falle die Lerner ein Phänomen auswählen und sich selbst erklären können (Kiel, 1999). Eng verbunden mit der Berücksichtigung des Lebensweltbezugs ist das Ziel, die Motivation der Schüler*innen zu steigern (Findeisen, 2017). Allerdings wird nach Helmke (2017) unter Unterrichtspraktikern kontrovers diskutiert, ob der Unterricht stets an der Lebenswelt der Schüler*innen orientiert sein soll und am Beispiel von Lebensweltbezug im Musikunterricht die „Musikvorlieben [*der Schüler*innen*, vom Autor hinzugefügt] nicht im Unterricht auch noch analysiert und zerredet werden“ (Helmke, 2017, S. 224; siehe dazu Rolle, 2011).

Fernab dieser Einschätzung begegnet das Lebenswelt-Konzept innerhalb der Musikpädagogik an verschiedenen Stellen, außerdem wird ihm eine große Bedeutung beigemessen (z. B. Ehrenforth, 1993 und 2001; Richter, 1993). In einer ausführlichen Zusammenfassung stellt Zimmer (2002) unter anderem fest, dass verschiedene Ansätze des Konzepts bestehen und in den musikpädagogischen Diskurs integriert wurden. Das deckt sich mit der kritischen Anmerkung von Schreiber (2005), dass über verschiedene Disziplinen hinweg nicht nur eine „allgemein akzeptierte Lebenswelt-Theorie“ fehle, sondern darüber hinaus auch „über den Lebensweltbegriff keine Einigkeit, nicht innerhalb der Wissenschaften und erst recht nicht in Bezug auf außerwissenschaftliche Verwendungen“ (S. 311–312) bestehe. Die vorliegende Arbeit folgt im Kontext von Musikunterricht dem Lebensweltbegriff nach Richter (1993), der auf musikalische Erfahrungen, Gewohnheiten, Vorlieben, Abneigungen und Kenntnisse der Schüler*innen abzielt.

Zusammenfassend kann die Adressatenorientierung als Leitgedanke guten Erklärens verstanden werden. Manifeste Aspekte sind in diesem Zusammenhang beispielsweise die Anknüpfung an das Vorwissen der Schüler*innen, das Herstellen eines Bezugs zu ihrem Alltag und ihrer Lebenswelt sowie Möglichkeiten der didaktischen Reduktion des erklärten Gegenstands durch die Lehrkraft. Im Kontext des Erklärens von Musiklehrkräften mangelt es diesbezüglich an musikpädagogischen Forschungsarbeiten. Allerdings spielen die Konzepte, die für den Aspekt Adressatenorientierung relevant sind, auch für Musikunterricht eine wichtige Rolle, weshalb eine ausführliche Auseinandersetzung in Kapitel 2.4 erfolgt.

Strukturiertheit

Ein weiterer Aspekt guten unterrichtlichen Erklärens ist der Grad an Strukturiertheit einer Erklärung (z. B. Findeisen, 2017; Schopf & Zwischenbrugger, 2015). Das Konzept Strukturiertheit ist gleichzeitig auch ein wichtiges Merkmal für Unterrichtsqualität (z. B. Brophy, 2000; Helmke, 2017) und wird von Jank und Meyer (2013) in Form eines Strukturmodells für den Musikunterricht nutzbar gemacht.

Die Strukturiertheit einer Erklärung kann an verschiedenen Stellen erhöht werden: Beispielsweise kann eine kurze Übersicht über den zu erklärenden Gegenstand am Beginn einer Erklärung Orientierung geben (Schopf & Zwischenbrugger, 2015; Wagner & Wörn, 2011) und die Struktur explizit beziehungsweise „nach außen erkennbar“ (Leisen, 2007, S. 461) machen. Möglichkeiten zur Umsetzung dessen können dabei sowohl Visualisierungen in Form eines *advanced organizer* (zur Lernwirksamkeit solcher Strukturierungshilfen siehe Helmke, 2017) oder eine Auflistung zentraler Fragen sein, die die Erklärung beantworten will (Schopf & Zwischenbrugger, 2015).

Weiterhin kann im Erklärprozess die schrittweise Einführung des zu erklärenden Inhalts zur Strukturiertheit beitragen (Pauli, 2015). Dabei ist es zunächst während der Vorbereitung einer Erklärung notwendig, einzelne Elemente des Gegenstands zu identifizieren und sinnvoll miteinander zu verbinden (Schopf & Zwischenbrugger, 2015). Für die Strukturierung einer Erklärung werden verschiedene Strategien vorgeschlagen: Sie kann chronologisch oder sachlogisch geordnet sein oder nach den Prinzipien *Problem und Lösung* oder *Ursache und Wirkung* vorgehen (Hargie, 2013; Schopf & Zwischenbrugger, 2015). Sprachlich können Konjunktionen, wie *weil* oder *um zu*, aber auch Präpositionen oder Adverbien die Erklärung anhand der Sachlogik des Themas strukturieren (Kiel, 1999). Auch kann der Logik des Lernprozesses gefolgt werden, indem beispielsweise vom Einfachen ausgegangen wird und mit jedem Schritt die Komplexität steigt (Schopf & Zwischenbrugger, 2015). Im Strukturmodell nach Jank und Meyer (2013) lassen sich diesbezüglich Gemeinsamkeiten feststellen, indem Musikunterricht anhand der Sachlogik von Inhalten sowie der Lernlogik von Schüler*innen strukturiert werden kann.

Gerade vor dem Hintergrund musikbezogenen Lernens weist Fuchs (2015) dem Erklären im Musikunterricht der Grundschule – wobei angenommen wird, dass das auch für den Sekundarstufenunterricht gilt – einen eindeutigen Platz zu: „Notation und theoretische Erklärungen sind sinnvoll und notwendig, wenn die Kinder über ausreichende musikalische Erfahrungen verfügen. Statt ohne Nachfrage Erklärungen zu geben, sollte die Lehrperson ihren Unterricht so anlegen, dass bei Kindern ein ‚Fragehunger‘ erzeugt wird“ (S. 102). Offen bleibt dabei einerseits, was unter einer *theoretischen Erklärung* zu verstehen ist und ob sich der Terminus beispielsweise auf eine rein verbale Erklärung bezieht. Andererseits ist auch unklar, ob die Ermöglichung musikalischer Erfahrungen nicht bereits Teil eines Erklär- beziehungsweise Vermittlungsprozesses ist oder sein kann. Jedoch geht aus der Aussage hervor, dass Musikkernen vom aktiven Handeln der Schüler*innen ausgeht und der Erwerb von Wissen zeitlich nachzuordnen ist beziehungsweise zuerst Musik erklingen soll, bevor sie in Notation dargestellt wird (Fuchs, 2015). Rechnet man diese akustischen wie optischen Elemente zum Erklärprozess dazu, ergibt sich hinsichtlich der Strukturiertheit beim Erklären im Musikunterricht ein deutlicheres Bild, das in Kapitel 2.4.1 ausführlich dargestellt wird.

Eine weitere wichtige Rolle spielt im fachübergreifenden Diskurs zur Strukturiertheit die Kohärenz einer Erklärung, also die inhaltliche Verbindung der einzelnen Erklärungsschritte (Pauli, 2015; siehe dazu auch Drollinger-Vetter, 2011). In diesem Zusammenhang kann Findeisen (2017) unter Bezugnahme auf mehrere Studien zeigen, dass erfahrene Lehrkräfte im Gegensatz zu angehenden besser in der Lage sind, eine kohärente

Erklärung zu geben. Am Ende einer Erklärung können knappe Zusammenfassungen – je nach Länge der Erklärung auch nach Zwischenabschnitten (Leisen, 2007) – dazu dienen, den Schüler*innen ein Gefühl der Vertrautheit zu vermitteln und die wichtigen Elemente der Erklärung hervorzuheben (Findeisen, 2017). Zur Strukturiertheit einer Erklärung kann außerdem beitragen, sich beim Erklären auf das Wesentliche des Gegenstandes zu fokussieren und ihn klar zu umreißen (Hargie, 2013; Pauli, 2015). Mögliche Ursachen, das Wesentliche bei einer Erklärung aus den Augen zu verlieren, sind beispielsweise Zwischenfragen von Schüler*innen, deren Beantwortung die Gefahr des Abschweifens birgt (Wagner & Wörn, 2011). Befunde aus dem Bereich des computerbezogenen beziehungsweise multimedialen Lernens weisen unter der Theorie *minimal instruction* darauf hin, dass Erklärungen nicht nur präzise und kohärent, sondern auch „mit möglichst wenig irrelevanten Details versehen sein sollten“ (Renkl et al., 2006, S. 212). Wenn es der Erklärgegenstand erfordert, sollten zudem prinzipienorientierte Erklärungen bevorzugt werden, in denen Regeln und Gesetzmäßigkeiten im Zentrum stehen (Renkl et al., 2006).

Bezüglich beider Aspekte, Kohärenz sowie Fokussierung auf das Wesentliche, lassen sich innerhalb der musikpädagogischen Forschung nur schwer Anknüpfungspunkte finden, um überhaupt Annahmen über die Rolle hinsichtlich des Erklärens im Musikunterricht treffen zu können. Außer der klaren Vorstellung über die Reihenfolge von Handlung vor Wissen beziehungsweise Klang vor Zeichen (Fuchs, 2015), die auch für das Erklären als Mikroform von Unterricht angenommen werden kann, adaptiert zwar auch Biegholdt (2013) die Merkmale guten Unterrichts nach Mayer (2004) und somit auch die „klare Strukturiertheit des Unterrichts“ (Biegholdt, 2013, S. 21) auf den Musikunterricht. Jedoch mangelt es dabei an einer konkreten Ausgestaltung, was das für den Musikunterricht im Allgemeinen oder übertragen auf das Erklären im Speziellen bedeutet. Während Strukturiertheit also zumindest implizit als Aspekt guten Musikunterrichts Erwähnung findet, ist hinsichtlich der Strukturiertheit als Qualitätsmerkmal unterrichtlichen Erklärens im Musikunterricht ein Desiderat festzustellen.

Sprachliche Verständlichkeit

Die Bildungsadministration deklariert eine adressatengerechte Darstellung fachlicher Inhalte in Verbindung mit einfacher Sprache als generelle fachdidaktische Kompetenz einer Lehrkraft (KMK, 2019). Diesbezüglich konnten Hattie et al. (2014) zeigen, dass die Verständlichkeit sprachlicher Äußerungen im Unterricht einen großen Effekt auf den Lernerfolg von Schüler*innen hat (Cohens $d = 0,75$). Helmke (2017) modelliert die Klarheit und Strukturiertheit von Unterricht als Qualitätsmerkmal und integriert dabei „Dimensionen der sprachlichen Verständlichkeit“ (S. 196). Dabei wird bereits deutlich, dass die sprachliche Verständlichkeit, ebenso wie die vorherigen Aspekte Adressatenorientierung und Strukturiertheit, zunächst einen allgemeinen und nicht genuin erklärerspezifischen Qualitätsaspekt darstellt.

Hinsichtlich unterrichtlichen Erklärens verweist jedoch auch Kiel (1999) auf die Verbindung zwischen sprachlicher Verständlichkeit sowie Strukturiertheit und fordert explizit die Verwendung spezieller Wortarten (z. B. Konjunktionen), um Erklärungen Struktur

und Kohärenz zu geben. Sprachliche Verständlichkeit ist auch bezüglich der Orientierung an Adressat*innen – weiterhin als Leitprinzip guten Erklärens verstanden – relevant, beispielsweise bei der Verwendung von Fachbegriffen beziehungsweise der Berücksichtigung von Vorwissen und Wortschatz der Schüler*innen (Wagner & Wörn, 2011). Dabei besteht fachübergreifend Einigkeit darüber, dass Sprache nicht nur in Fächern wie Mathematik, Deutsch und Chemie (z. B. Becker-Mrotzek et al., 2013; Leiss et al., 2017; Sumfleth et al., 2013), sondern auch im Fach Musik eine wichtige Rolle spielen sollte, jedoch „aus musikpädagogischer Perspektive weder hinreichend theoretisch noch fachdidaktisch aufgearbeitet zu sein“ (Ahlers & Seifert, 2015, S. 235) scheint. Auch Rolle (2014) sieht für die Frage, wie im Musikunterricht über Musik gesprochen wird, erheblichen musikpädagogischen Forschungsbedarf, der mittlerweile in ersten Arbeiten jedoch eingelöst wird (z. B. Barth & Bubinger, 2015; Bossen, 2019; Bossen & Jank, 2017).

Bei der Sichtung fachübergreifender didaktischer Literatur wird an vielen Stellen deutlich, dass das Verhältnis von sprachlichen und sprecherischen Aspekten guten Erklärens unterschiedlich gehandhabt wird: Auf der einen Seite werden sowohl im Kontext der Unterrichtsforschung (z. B. Helmke, 2017) als auch hinsichtlich der fachdidaktischen Literatur zum Erklären von Lehrkräften (z. B. Findeisen, 2017; Wagner & Wörn, 2011) unter dem Begriff Sprache nicht nur sprachliche, sondern auch sprecherische Aspekte, wie Vortragsweise, Stimmklang oder Körpereinsatz der Lehrkraft, subsummiert – wobei in Bezug auf die Arbeit von Findeisen (2017) jedoch einzuschränken ist, dass die Autorin sprachliche Aspekte in ihrer Untersuchung außen vorlässt und dies damit begründet, dass Sprache in der simulierten Erklärung „aufgrund der konstruierten Situation nicht authentisch beurteilt werden“ (S. 180) könne. Auf der anderen Seite trennen beispielsweise Schopf und Zwischenbrugger (2015) klar zwischen sprachlichen und sprecherischen Aspekten und stellen heraus, dass sich Merkmale des Sprech- und Körperausdrucks dem Handeln der Lehrkraft zuordnen lassen und die sprachlichen Aspekte eher auf konzeptueller Ebene der Erklärung anzusiedeln sind. Die vorliegende Arbeit folgt dieser Unterteilung, weshalb sprecherische Aspekte im weiteren Verlauf noch gesondert dargestellt werden.

Es finden sich viele Anmerkungen darüber, welche Rolle der Sprache beim Erklären zukommen sollte. In einer allgemeinen Formulierung fordert Leisen (2007), „Erklärungen müssen in eine passende Sprache gefasst werden“ (S. 461). Während in diesem Zitat die Orientierung beziehungsweise Passung an die Adressat*innen deutlich wird, bleibt jedoch offen, unter welchen Bedingungen Sprache passend ist. Auch vermeintlich konkretere Hinweise zum unterrichtlichen Erklären wie „die Sprache beim Lehrervortrag ist kurz, direkt und präzise [*sic*]“ (Lehner, 2018, S. 157; siehe auch Helmke, 2017) lassen Raum für Interpretationen. Aus sprachwissenschaftlicher Perspektive bewerten Thim-Mabrey und Lindl (i. Vorb.) solche Ratschläge als unzeitgemäß für heutige Stilpräferenzen, die zudem an Empfehlungen zur Verknüpfung und Verdichtung nicht vorbeikommen, wie sie Lehner (2018) im weiteren Verlauf entgegen seiner Aussage zur Sprache beim Lehrervortrag fordert. Die Autorin und der Autor entfalten die Rolle sprachbezogener Aspekte beim unterrichtlichen Erklären auf Wort-, Satz- sowie Textebene und lassen

dabei sprachwissenschaftliche Erkenntnisse der Verständlichkeits- und der Mündlichkeitsforschung einfließen (siehe ausführlich Thim-Mabrey & Lindl, i. Vorb.).

Dabei gelten mündliche Erklärungen als Texte, die mit sprachwissenschaftlichem Instrumentarium auf ausgewählte wesentliche Eigenschaften hinsichtlich ihrer Verständlichkeit untersucht werden können (Thim-Mabrey & Lindl, i. Vorb.). Weitere wesentliche Aspekte sind die sprachliche Realisation unterrichtlichen Erklärens als Form mündlicher Kommunikation sowie die Aufbereitung der Fachinhalte für die Zielgruppe der Schüler*innen (Thim-Mabrey & Lindl, i. Vorb.), also die Orientierung an den Adressat*innen unterrichtlichen Erklärens.

Auf der Wortebene wird beim unterrichtlichen Erklären die angemessene Verwendung von Fachbegriffen mit großer Übereinstimmung als wichtig erachtet (z. B. Aeschbacher, 2009; Drollinger-Vetter, 2011; Findeisen, 2017; Kulgemeyer & Tomczyszyn, 2015; Wagner & Wörn, 2011). Drollinger-Vetter (2011) betont die dafür notwendige Aufmerksamkeit der Lehrkräfte, bei Bedarf die jeweils gebrauchten Begriffe – gerade wegen der häufig ganz selbstverständlichen Verwendung im eigenen Sprachgebrauch – auf ihre zentralen Elemente hin aufzufalten und für Lernende verständlich zu machen (siehe auch hier die Diskussion zum expert blind spot, z. B. Renkl et al., 2006). Im Kontext von Musikunterricht erhebt Rolle (2014) den Vorwurf, dass dort „zwar grundlegende musiktheoretische Fachbegriffe gelernt“ werden, dies jedoch „überwiegend schematisch, ohne lebendige Verwendungskontexte“ (S. 3) vollzogen wird.

Auf Satzebene sind „nicht einfach die ‚Kürze‘, sondern die Konstruktionseigenschaften von Sätzen“ relevant (Bredel & Maaß, 2016, S. 384). Gerade Nebensatzkonstruktionen müssen dabei nicht nur in schriftlichen, sondern auch in mündlichen Kommunikationen mit Bedacht eingesetzt werden (Thim-Mabrey & Lindl, i. Vorb.). Folgt eine unterrichtliche Erklärung dem Prinzip der leichten Sprache⁴¹, sind Nebensätze strikt zu vermeiden sowie die Verknüpfungen der (Haupt-)Sätze und die Stellung von Worten innerhalb eines Satzes zu beachten (Bredel & Maas, 2016).

Auf Textebene – die mündlichen Erklärungen werden wie eingangs beschrieben als Text aufgefasst – kann die Struktur eines Textes durch sprachliche Elemente verdeutlicht werden und zur Verständlichkeit beitragen (Thim-Mabrey & Lindl, i. Vorb.). Beispielsweise gliedern Aufzählungswörter wie *erstens*, *zweitens*, *drittens* die Textstruktur oder sprachliche Hinweise wie *Das ist jetzt sehr wichtig* heben besondere Stellen der Erklärung hervor (Thim-Mabrey & Lindl, i. Vorb.). Im Gegensatz zum Umgang mit Fachbegriffen (Wortebene), die im musikpädagogischen Kontext eine ebenso wichtige Rolle wie die vertraute Anwendung von Fachsprache durch die Schüler*innen spielen (Rolle, 2014), sind Satz- und Textebene vor allem als linguistische Kategorien zu sehen, die jedoch generell zur sprachlichen Verständlichkeit einer Erklärung beitragen können.

Die Sprache von Lehrkräften bewegt sich beim unterrichtlichen Erklären im Spannungsfeld zwischen Aspekten mündlicher Sprachlichkeit, die beispielsweise verständlich sein soll und sich eher an der Alltagssprache der Schüler*innen orientiert, und Aspekten

41 Durch die Einhaltung der Prinzipien Leichter Sprache sollen umfassende Partizipationsmöglichkeiten durch möglichst für alle Personen verständliche Texte gewährleistet werden. Für eine vertiefte Auseinandersetzung siehe Bredel und Maas (2016).

von Bildungssprachlichkeit, wodurch bei Schüler*innen Sprachkompetenz anhand der Lehrkraft als sprachliches Vorbild gefördert werden soll (Thim-Mabrey & Lindl, i. Vorb.; siehe auch Kulgemeyer & Schecker, 2013). Eine Lehrkraft sollte also nicht nur die Fachsprache beherrschen, sondern auch das Sprachniveau und die Sprachpraxis seiner Schüler*innen kennen, um instruktionale Erklärungen erfolgreich geben zu können (Leinhardt, 2001). Bei der Abwägung zwischen dem Anteil an Alltags- und Fachsprache sollte die Prämisse fachlicher Korrektheit jedoch nie durch „sprachliche Unschärfen oder eine didaktische Reduktion“ (Wagner & Wörn, 2011, S. 29) verletzt werden.

Aus musikpädagogischer Perspektive rekurriert Rolle (2014) diesbezüglich auf verschiedene Ebenen des Sprechens über Musik – gemeint ist die verwendete Sprache –, wie sie Reinecke (1974) skizziert hat, und ruft sie für einen musikpädagogischen Diskurs über Fachsprache in Erinnerung. Unterteilt wird dabei zwischen einer umgangssprachlichen (Laiensprache), musikpraktischen (Musiker*innensprache) und musikwissenschaftlichen (objektive Fachsprache) Verständigung über Musik, deren Hierarchisierung für Rolle (2014) jedoch zweifelhaft erscheint. Die verschiedenen Sprachebenen können auch für das Erklären im Musikunterricht als Orientierung dienen, wobei diesbezüglich gerade ein Kontinuum zwischen Laien- und Fachsprache als sinnvoller Rahmen erscheint, da nach Rolle (2014) gerade die Position der Musiker*innensprache vage bleibt.

Die Relevanz von Sprache und sprachlicher Verständlichkeit in allen Fächern hebt Schmölzer-Eibinger (2013) hervor, da Sprache nicht nur das primäre Medium zur Vermittlung von Wissen, sondern auch des Lernens sei. Ein Wissenserwerb im Fachunterricht ist ohne Sprache so gut wie nicht möglich (Schmölzer-Eibinger, 2013; Wellenreuther, 2013). Die Autorin fordert daher für alle Lehrkräfte, einen „sprachbewussten Fachunterricht zu realisieren“ (Schmölzer-Eibinger, 2013, S. 36).

In der Musikpädagogik setzt sich unter anderem Pendorf (2019) mit sprachsensiblen Musikunterricht auseinander und sieht für die Musikpädagogik als Aufgabe, „im Einklang mit der Debatte um das Phänomen der Bildungs(un)gerechtigkeit“ Handreichungen zur Verfügung zu stellen, „die jede Musiklehrkraft dazu befähigt, ihren Unterricht sprachsensibel zu gestalten“ (S. 13). Ahlers und Seifert (2015) entwickelten ein Testinstrument, das die Anwendung von Fachsprache durch Schüler*innen anhand präparierter Fachtexte messen kann. Indem Wortendungen getilgt wurden, konnte sowohl „ein hohes Maß an allgemein-sprachlichen Fähigkeiten“ als auch „spezifischer, fachsprachlicher Parameter“ (Ahlers & Seifert, 2015, S. 240) erfasst werden. Im Rahmen ihrer Studie konnten die Autoren zeigen, dass Schulbuchtexte „für viele Schüler*innen dieser Stichprobe schwer zu verstehen“ (Ahlers & Seifert, 2015, S. 245) sind, und führen dies anhand der Aussagen von Musiklehrkräften unter anderem auf eine Fokussierung hin zu vermehrter Musikpraxis zurück. Aufgrund dieser Ergebnisse ist zu erwarten, dass die sprachliche Verständlichkeit von Erklärungen im Kontext von Musikunterricht eine wichtige Rolle spielt.

Sprech- und Körperausdruck

Da Lehrkräfte beim Unterrichten viel sprechen und ihre Stimme einsetzen, zählen sie Rogerson und Dood (2005) zur Gruppe der „professional voice users“ (S. 47). Vor diesem

Hintergrund wird in Deutschland einerseits gefordert, die stimmliche Ausbildung im Lehramtsstudium flächendeckend auszubauen (z. B. Lemke, 2006; Lukaschyk, 2018; Nusseck et al., 2020).

Innerhalb der Ausbildung von Musiklehrkräften spielen derartige Seminarkonzeptionen eine Rolle, die beispielsweise das Singen, Sprechen und Präsentieren beim Unterrichten thematisieren, da gerade der Wechsel zwischen Gesangs- und Sprechstimme für Musiklehrkräfte eine zusätzliche Herausforderung bereithält (Fröhlich-Schweiger, 2019; Pezenburg, 2018). Die Entwicklung der (Sing-)Stimme spielt beispielsweise auch im Kontext von Fort- und Weiterbildungen fachfremd unterrichtender Musiklehrkräfte an Grundschulen eine wichtige Rolle (Forge, 2016). Gerade in Bezug auf das stimmliche Selbstkonzept der Lehrenden (siehe dazu z. B. Niessen, 2007) können hier positive Effekte nachgewiesen werden, die sich beispielsweise in einer gesteigerten Selbstwahrnehmung der eigenen Singfähigkeit widerspiegeln (Forge, 2016). Andererseits wird auch diskutiert, die stimmliche Eignung angehender Lehramtsstudierender bereits vor Beginn des Studiums – dies ist beispielsweise für einige Musik-Lehramtsstudiengänge in Bayern⁴² oder für Ausbildungen in anderen sprechintensiven Berufsfeldern wie Schauspiel oder Logopädie üblich – zu überprüfen, da „stimmliche und sprecherische Auffälligkeiten bei Lehramtsanwärtern in unangemessen hoher Zahl auftreten“ (Lemke, 2006, S. 27). Lemke (2006) begründet diesen Befund zum einen mit der fehlenden Begutachtung zu Beginn des Studiums und zum anderen mit der als Folge daraus ausbleibenden Therapie.

In Bezug auf das unterrichtliche Erklären unternehmen Gunga et al. (i. Vorb.) erste sprechwissenschaftliche Forschungsbemühungen, da bisher vor allem fachdidaktische wie kommunikationswissenschaftliche Beiträge (z. B. Hargie, 2013; Odora, 2014; Schopf & Zwischenbrugger, 2015) die Rolle des Sprechens einer Lehrkraft im Unterricht thematisieren und dabei auch Bezüge zum unterrichtlichen Erklären herstellen. Neben physiologischen Anforderungen an die Stimme einer Lehrkraft werden aus sprecherischer Perspektive darüber hinaus auch rhetorische Fähigkeiten gefordert, um beispielsweise durch den Stimmeinsatz bei Schüler*innen „Aufmerksamkeit zu erzeugen“ (Burger, 2018, S. 17; siehe auch Odora, 2014). Darüber hinaus fassen Rogerson und Dodd (2005) auf Basis mehrerer empirischer Studien zusammen, dass sich Stimmstörungen bei Lehrkräften negativ auf das Lernen von Schülerinnen und Schülern auswirken.

Welche anderen Effekte durch rhetorische Qualität erzielt werden können, zeigt eine Studie, die zwar keine schulische Unterrichtssituation untersucht, aber aus einem Lehrkontext stammt. Abrami et al. (1982) konnten zeigen, welche positiv-trägerischen Einfluss das Sprechen nehmen kann, indem selbstbewusste Rhetorik über inhaltliche Fehler im Vortrag hinwegtäuschen können (*Dr.-Fox-Effekt*; siehe auch Helmke, 2017). Übertragen auf das unterrichtliche Erklären kann die rhetorische Fähigkeit einer Lehrkraft als performatives Element unterrichtlichen Erklärens in zweifacher Hinsicht eine wichtige Rolle spielen. Zum einen wurde bereits ausgeführt, dass die Vortragsweise der Lehrkräfte positive Effekte, wie beispielsweise die Erzeugung von Aufmerksamkeit bei Schüler*innen,

42 Bei einigen bayerischen Aufnahmeprüfungen zum Studium des Fachs Musik für das Lehramt Gymnasium oder für Musik als Unterrichtsfach in Grund-, Mittel- und Realschule ist nicht nur Vorsingen vorgesehen, sondern es muss auch ein Sprechstück vorgetragen werden.

erzielen kann. Zum anderen birgt eine sprecherisch überzeugende Performanz jedoch die Gefahr, beispielsweise Mängel hinsichtlich der fachlichen Korrektheit der Erklärung zu überdecken (Schopf & Zwischenbrugger, 2015). In Bezug auf die sprecherische Performanz von Musiklehrkräften sind keine empirischen Studien bekannt.

In der Sprechwissenschaft wird eine Kategorisierung vorgenommen, die das Sprechen in verbale, paraverbale (z. B. Sprechgeschwindigkeit) und extraverbale (z. B. Blickverhalten) Aspekte unterteilt (Heilmann, 2002). Neben verbalen Aspekten, die bereits im vorherigen Abschnitt ausführlicher herausgearbeitet wurden (siehe Sprachliche Verständlichkeit), sind mehrere parasprachliche Aspekte, wie beispielsweise der Stimmklang, für unterrichtliches Erklären relevant (Schopf & Zwischenbrugger, 2015). Befunde zum Stimmklang von Lehrkräften zeigen nicht nur, dass sich Stimmstörungen wie eine heisere Stimme nicht auf das Verhalten von Schüler*innen in Form von Aufmerksamkeits- sowie Disziplinproblemen, sondern auch auf deren Verstehens- und Lernleistung negativ auswirken können (Voigt-Zimmermann, 2017). Schopf und Zwischenbrugger (2015) nennen neben einer deutlichen Aussprache (Artikulation) für das Erklären einer Lehrkraft die Relevanz weiterer wichtiger Parameter wie Sprechtempo, Verwendung von Sprechpausen, Lautstärke und Variation. Beispielsweise kann eine überlegte Pausensetzung bei den Adressat*innen zum besseren Verstehen beitragen, während dies bei einer undeutlichen Aussprache der Lehrkraft grundsätzlich nicht gegeben sein kann (Hargie, 2013). Die bisher genannten prosodischen Merkmale werden in Bezug auf generelle sprecherische Fähigkeiten in der Fachliteratur ausführlich dargestellt und um weitere Parameter wie die Sprechmelodie, Sprechspannung oder das Timbre der Stimme ergänzt (z. B. Allhoff & Allhoff, 2010; Bose et al., 2016; Kranich, 2016).

Unter extraverbalen Aspekten sprechsprachlicher Kommunikation fasst Heilmann (2002) beispielsweise Mimik, Gestik und Körperausdruck. Mit Blick auf das Erklären in Unterrichtssituationen ergänzen Schopf und Zwischenbrugger (2015) diese Formen nonverbaler Kommunikation noch um weitere Aspekte wie Blickverhalten der Lehrkraft oder die vorhandene Raumregie. Nach Meinung der Autorinnen könnten sich beispielsweise zu häufige Blicke auf eigene Unterlagen beziehungsweise Medien oder eine einseitige räumliche Positionierung beim Unterrichten beziehungsweise Erklären als problematisch für die Aufmerksamkeit der Schüler*innen erweisen (Schopf & Zwischenbrugger, 2015). In positiver Weisen kann Körpersprache jedoch auch dazu beitragen, besonders wichtige Aspekte innerhalb der Erklärung hervorzuheben (Findeisen, 2017). Beispielsweise können die Hände, der Kopf oder auch die Augen zweckgerichtet eingesetzt werden, um wichtige inhaltliche Stellen zu betonen (Brown & Hatton, 1982; Hargie, 2013). Gunga et al. (i. Vorb.) merken an, dass die Sprechwissenschaft die Untersuchung von Körpersprache einer Lehrkraft im Kontext von Unterricht bisher nicht ausreichend abdecken konnte, weshalb auf Erkenntnisse der *teacher-enthusiasm*- und *teacher-immediacy*-Forschung zurückgegriffen werden sollte. Die *teacher-enthusiasm*-Forschung untersucht den Einfluss von Verhaltensaspekten von Lehrkräften „wie ausgeprägte Gestik, wechselnde Intonation, ständiger Blickkontakt, häufige Standortwechsel auf der ‚Bühne‘ sowie Humor“ hinsichtlich „der Frage nach dem Unterrichtserfolg“ (Helmke, 2017, S. 114).

Demgegenüber verfolgt die *teacher-immediacy*-Forschung das Ziel, Verhaltensweisen von Lehrkräften daraufhin zu untersuchen, ob sie die physische wie psychische Distanz zu Schüler*innen reduzieren können (Andersen, 1979). Gemeinsam ist beiden Forschungssträngen, dass sie vergleichbare manifeste Merkmale in den Blick nehmen (Gunga et al., i. Vorb.). Für das Unterrichten im Fach Musik zeigt sich diesbezüglich gerade der Einsatz von Gestik beim gemeinsamen Musizieren für die Lehrkraft als relevant. Sowohl beim Dirigieren als auch bei chorischer Stimmbildung können bestimmte Bewegungsmuster mit den Händen stimmtechnische Einstellungen abbilden oder durch einen harten Schlag beziehungsweise weiche fließende Bewegungen unterschiedliche Singweisen befördert werden (Pezenburg, 2018; siehe auch Hempel, 2008).⁴³ In ihrem Modell professioneller Dispositionen von Musiklehrkräften setzt Puffer (2021) musikpraktisch-künstlerische Fähigkeiten und Fertigkeiten an eine zentrale Stelle, in der affektiv-motivationale, kognitive, perzeptive und sensomotorische Komponenten zusammenfließen.

Zusammenfassend kann festgestellt werden, dass die Stimme, die Sprechweise und der Körperausdruck von Lehrkräften in Zusammenhang mit Erklären als Sprechhandlung im Unterricht bis dato kaum eine Rolle spielen, denn es fehlt nicht nur an Modellen, sondern mit wenigen Ausnahmen (z. B. Schopf & Zwischenbrugger, 2015) auch an empirischen Erkenntnissen. Da Studien zur Sprechwirksamkeitsforschung jedoch zeigen konnten, dass sprecherische Merkmale in kommunikativen Kontexten wirksam sind, sollte auch beim unterrichtlichen Erklären der performative Akt des Sprechens in den Blick genommen werden (Gunga et al., i. Vorb.). Dies gilt auch für den Unterricht im Fach Musik, da es auch innerhalb der musikpädagogischen Fachliteratur hinsichtlich des Einsatzes sprecherischer Fähigkeiten an Forschungsbeiträgen mangelt.

Einsatz von Repräsentationen

Der Einsatz von Visualisierungen, Beispielen, Analogien oder Metaphern wird mit großer Übereinstimmung als wichtiges Merkmal guten unterrichtlichen Erklärens genannt und häufig unter den Begriffen *Repräsentation* beziehungsweise *Repräsentationsform* zusammengefasst (Findeisen, 2017; siehe z. B. Geelan, 2012; Kulgemeyer & Tomczyszyn, 2015; Pauli, 2015). Schnotz und Bannert (1999) definieren⁴⁴ eine Repräsentation als „ein Objekt oder ein Ereignis, das für etwas anderes steht, es re-präsentiert“ (S. 5) und unterscheiden dabei zwischen externen Formen der Repräsentation (z. B. durch Analogien, Texte oder Bilder) und internen (mental) Repräsentationen. Beide Repräsentationen können miteinander interagieren, indem externe Repräsentationen eine Person zur Bildung interner mentaler Repräsentationen anregen können (Schnotz & Bannert, 1999). Für das unterrichtliche Erklären kann demzufolge angenommen werden, dass die Verwendung externer Repräsentationen in Form von Visualisierungen oder Analogien zum

43 Diese Form der Sichtbarmachung beziehungsweise Visualisierung von Musik wird unter dem folgenden Aspekt Repräsentationsformen noch einmal aufgegriffen.

44 Sie rekurren in ihrer Definition auf Peterson (1996).

Verstehen in Form der Bildung oder Weiterentwicklung mentaler Repräsentationen führen kann (Hartig, 2015).

Interne mentale Repräsentationen können nach Hartig (2015) als „Abbildung von Wissensstrukturen im Kopf eines Individuums“ verstanden werden, wobei sie kein detailgetreues Ebenbild externer Repräsentationen darstellen; gespeichert werden „nur bedeutsame Merkmale eines extern repräsentierten Sachverhalts“, während auf „unwichtige Details“ (S. 13) verzichtet wird. Die Autorin nennt als Beispiel, dass das Konzept eines Autos durch Merkmale wie vier Räder oder Lenkrad repräsentiert wird (Hartig, 2015). Aus musikpädagogischer Sicht spricht Gruhn (1994) von musikalischen Repräsentationen und meint damit ebenfalls solche mentalen Vorgänge, bei denen man sich zum Beispiel „bekannte Melodien innerlich vorstellen“ oder „angefangene Melodien weiterdenken“ (S. 13) kann. Eine solche Verwendung des Begriffs Repräsentation ist im Kontext von Qualitätsmerkmalen unterrichtlichen Erklärens nicht gemeint.

Externe Repräsentationsformen beziehen sich im Gegensatz dazu vor allem auf optische⁴⁵ und verbale Darstellungsformen (z. B. Findeisen, 2017; Hill et al., 2012). Shulmans (1986) Ausführungen, bei denen „ways of representing and formulating the subject“ (S. 9) nicht nur Illustrationen, sondern auch Analogien und Beispiele meinen, werden in der Konzeptualisierung der COACTIV-Studie mit der Terminologie „Wissen über multiple Repräsentations- und Erklärungsmöglichkeiten“ (Baumert & Kunter, 2011, S. 38) zusammengefasst. Allerdings bleibt an dieser Stelle unklar, in welchem Verhältnis Repräsentieren und Erklären im COACTIV-Modell zueinanderstehen. Eine andere Konzeptualisierung, die ebenfalls aus dem Bereich Mathematik stammt, definiert eine Repräsentation typischerweise als optische oder verbale Darstellung quantitativer Informationen (Hill et al., 2012).

Im Zusammenhang mit der vorliegenden Arbeit liegt der Fokus auf externen Repräsentationsformen, die beim Erklären zum Einsatz kommen und (musikalische) Phänomene in verschiedener Form abbilden und veranschaulichen können. Unter Rekurs auf die Kategorisierung von Hill et al. (2012) identifiziert Findeisen (2017) bei einer Vielzahl an Studien zur Qualität guten Erklärens fünf externe Formen der Repräsentation, die der folgenden Darstellung als Orientierung dienen und dabei in vier zentralen Aspekten dargestellt werden:

- (1) Visualisierung der Inhalte,
- (2) Verwendung von Beispielen,
- (3) Verwendung von Analogien und
- (4) Verbindung verschiedener Repräsentationsformen.

45 Die vorliegende Arbeit verwendet für externe Formen der Repräsentation die Begriffe *verbal*, *optisch* und *akustisch*. In anderen Arbeiten wird statt *optisch* vorzugsweise der Begriff *visuell* verwendet (z. B. Findeisen, 2017) – es wird vermutet, dass dies unter anderem aufgrund der Übersetzung aus dem Englischen vollzogen wird. In der hier zugrundeliegenden begrifflichen Rahmung steht *visuell* mit Wahrnehmungsprozessen und demnach internen Formen der Repräsentation in Verbindung. Mit *optisch* werden externe Reize, wie beispielsweise farbige Markierungen, bezeichnet (siehe für eine solche Verwendung der Begriffe *optisch* und *akustisch* z. B. Rösler, 2011, S. 14).

(1) Visualisierungen

Schopf und Zwischenbrugger (2015) betrachten bereits das Verschriftliche zentrale Begriffe oder Merksätze an der Tafel beziehungsweise dem Whiteboard als Visualisierung, wobei die Autorinnen vor allem solche Visualisierungen als sinnvoll erachten, die zu mentalen Repräsentationen bei Schüler*innen führen. Dementsprechend zählt Findeisen (2017) vor allem Visualisierungen, die Erklärungsinhalte abbilden und grafisch repräsentieren, zu Qualitätsaspekten guten Erklärens, da sie „eine Verbindung der verbalen Informationen mit bildhaften Darstellungen“ (S. 69) ermöglichen. Auch Hargie (2013) betont diesbezüglich, dass Menschen Informationen über unterschiedliche sensorische Kanäle aufnehmen. Er plädiert deshalb für einen multimedialen Ansatz, der beispielsweise gesprochenes Wort und optische Darbietung kombiniert.

Beim unterrichtlichen Erklären geht es darum, Inhalte durch den Einsatz von Visualisierungen (z. B. grafische Darstellungen, Tabellen, Zeichnungen, Concept Maps) besser verständlich zu machen (Pauli, 2015). Konkret kann eine Lehrkraft beim unterrichtlichen Erklären mittels Visualisierungen relevante Aspekte hervorheben (Findeisen, 2017) oder „die zentralen Elemente und Beziehungen eines Begriffs unmittelbar ablesbar machen“ (Pauli, 2015, S. 45). Trotz der Beobachtung eines steigenden Forschungsinteresses stellt Geelan (2012) ein Desiderat bezüglich des Einsatzes von Visualisierungen beim Erklären in Lehr-Lernkontexten fest und verweist darauf, dass dieser Aspekt eher in der allgemeinen Diskussion über unterrichtliches Erklären verhandelt wird. In dieser Weise untersuchen Kulgemeyer und Tomczyszyn (2015) in ihrer Studie zur Erklärens-fähigkeit von Physiklehrkräften auch den Einsatz graphischer Darstellungen. Findeisen (2017) konnte für das Erklären im Bereich Rechnungswesen zeigen, dass die meisten der angehenden Lehrkräfte in den simulierten Erklär-situationen eine grafische Darstellung nutzen.

Im Kontext von Erklären im Musikunterricht meint die Visualisierung von Inhalten nicht nur die Notation beziehungsweise Notenschrift als Visualisierung klingender Musik, wie sie beispielsweise Bruhn (2019) als mediale Darstellung oder Biegholdt (2013) im Kontext von Unterricht beschreiben. Sie versteht sich hier vor allem auch als Form der grafischen Darstellung eines Inhalts, wie die vertikale Abbildung von Tondauer auf einem eindimensionalen Zeitstrahl oder das (Nach-)Zeichnen von Melodieverläufen mittels grafischer Linienführung (Krämer, 2007). Jünger (2006) konnte in einer Befragung von Musiklehrkräften über die Nutzung von Schulbüchern im Unterricht beobachten, dass Lehrkräfte Lehrbücher verwenden, um „verbale Informationen durch Abbildungen zu veranschaulichen“ (S. 214). Unklar bleibt dabei allerdings, wie diese Abbildungen beschaffen sein sollen.

Harnischmacher (2002) untersuchte bei Schüler*innen der Primarstufe den Einfluss computergestützter Visualisierungen auf die Melodiewahrnehmung und das Benennen von Tonhöhenrichtungen. Als Visualisierungen verstand er dabei unter anderem Darstellungen, in denen anhand vertikaler Achsen Unterschiede in der Tonhöhe abgebildet wurden, wobei in der Untersuchung bezüglich des generellen Einsatzes von Visualisierungen keine signifikanten Effekte auf die Melodiewahrnehmung der Schüler*innen festgestellt wurden (Harnischmacher, 2002). Hempel (2008) unterscheidet mehrere Formen der Visualisierung von Musik und fasst darunter neben der Notation von Musik auch Formen

des Dirigats, die musikalische Parameter durch Gesten visualisieren können. Unter musikdidaktischen Aspekten zählt der Autor beispielsweise „die Darstellungsmethode einfacher Liedformen mit Grafikelementen und Buchstaben“ (Hempel, 2008, S. 198). Über die Rolle von Visualisierungen beim Erklären im Musikunterricht liegen bislang keine empirischen Erkenntnisse vor.

(2) Beispiele

Die Verwendung von Beispielen spielt eine wichtige Rolle im Kontext unterrichtlichen Erklärens und wird an vielen Stellen als Aspekt guten Erklärens angeführt (z. B. Brown & Hatton, 1982; Findeisen, 2017; Leinhardt, 2001; Schopf & Zwischenbrugger, 2015). Nach Findeisen (2017) helfen Beispiele, „den Erklärungsgegenstand zu veranschaulichen und einzuordnen“, indem sie beispielsweise „die Gesetzmäßigkeiten eines Sachverhalts bzw. die Grenzen eines Konzepts oder auch Anwendungsmöglichkeiten des Sachverhalts verdeutlichen“ (S. 67). Während sowohl Brown und Hatton (1982) als auch Leinhardt (2001) für eine sorgfältige Planung und Auswahl an Beispielen plädieren, die zwar nicht alle zum Einsatz kommen müssen, aber bei der Lehrperson parat sein sollten, ergänzt Findeisen (2017) mehrere Anforderungen an ein gutes Beispiel, etwa dass es in „Bezug zur Lebenswelt der Lernenden“ (S. 68) steht. Schopf und Zwischenbrugger (2015) geben Empfehlungen darüber ab, wie Beispiele eine Erklärung im Wirtschaftskontext unterstützen können: Diese sollen (a) passend und für den erklärten Gegenstand typisch, (b) deckend beziehungsweise durchgängig nutzbar, (c) konkret, (d) realitätsnah beziehungsweise realistisch und (e) schülernah sein. Für den erst genannten Aspekt *passend/typisch* konkretisiert Kiel (1999): „Das typische Beispiel hat Vorrang vor einem Gegenbeispiel oder einem weniger typischen Beispiel, weil die Präsentation des Nichttypischen bei der Einführung eines neuen Begriffs leicht zu Fehlkonzeptionen führen kann oder zumindest die Denkkaktivität der Lerner erschwert“ (S. 287).

Für den Musikunterricht wird angenommen, dass grundsätzlich alle Attribute übertragbar sind, wenngleich *konkret* und *realitätsnah* von Schopf und Zwischenbrugger (2015) in ihren Ausführungen vor allem fachspezifisch ausgeleuchtet werden. Bei der Vermittlung musikalischer Inhalte können Beispiele auf verschiedene Art und Weise eingesetzt werden. Das Verhältnis der binären Unterteilung zweier Notenwerte kann exemplarisch an der Tondauer einer Ganzen sowie zweier Halber Noten verdeutlicht werden. Verschiedene Möglichkeiten der Punktierung können innerhalb eines bestimmten Metrums beispielsweise anhand ausgewählter Rhythmuspatterns präsentiert werden, die als (Noten-)Beispiele dienen und unterschiedliche punktierte Notenwerte bereithalten. Auch können Werke von Komponisten wie Boulez oder Stockhausen als konkrete Beispiele dienen, um das Kompositionsverfahren der Aleatorik zu veranschaulichen. Während es bei Visualisierungen eher um die optische Darstellung eines Beispiels geht, kann dieses inhaltlich verschieden gestaltet werden und unterschiedliche Umfänge annehmen.

Generell gibt es für die Verwendung von Beispielen innerhalb einer Erklärung zwei Prinzipien: Beim induktiven Vorgehen wird das Grundprinzip aus einem oder mehreren Beispielen herausgearbeitet, während das deduktive Vorgehen zunächst das Grundprinzip erläutert und anhand von Beispielen illustriert (Schopf & Zwischenbrugger, 2015).

Diesbezüglich konnten Reid und Roberts (2004) anhand einer Studie mit 82 Schüler*innen für den Mathematikunterricht beobachten, dass die Lernenden Erklärungen mit mehreren Beispielen und einem induktiven Vorgehen gegenüber deduktiven Erklärungen bevorzugen. Inwieweit beide Vorgehensweisen bei der Vermittlung von musikalischen Inhalten wirksam sind, wurde aus musikpädagogischer Perspektive noch nicht untersucht.

(3) Analogien

Auch der Einsatz von Analogien wird an verschiedenen Stellen als Merkmal guten Erklärens genannt (z. B. Findeisen, 2017; Geelan, 2012; Schopf & Zwischenbrugger, 2015). Dabei wird nicht immer explizit differenziert, wie diese Analogien repräsentiert werden. Beispielsweise liegen auch Visualisierungen Analogien zu Grunde, wenn eine Treppe als Hilfestellung für das Konzept der stufenweisen Reihung von Tönen in einer diatonischen Tonleiter fungiert. Hier sind insbesondere Analogien gemeint, die in einer Erklärung genannt werden, ohne dass dazu andere Formen der Repräsentation vorliegen. Vor diesem Hintergrund treten auch Konzepte wie *Metapher* oder *Vergleich* auf (z. B. Lehner, 2018; Neumeister, 2011). Die Fokussierung auf diese drei Begriffe liegt in der mehrmaligen Nennung in Bezug auf Merkmale guten Erklärens begründet und geschieht vor dem Wissen um weitere (nahe) verwandte Konzepte wie Gleichnis oder Symbol (Kohl, 2007).⁴⁶

Alle drei Begriffe unterscheiden sich vom *Beispiel*, das „als Teil des Erklärten anzusehen ist“, indem bei ihnen die „Differenz zwischen dem Erklärenden und dem Erklärten zentral ist“ (Kohl, 2007, S. 76). Nach Neumeister (2011) ist einer Metapher eine „(metaphorische) Bedeutungsverschiebung“ zentral, „da bei der Interpretation das Konzept eines Wortes auf einen anderen Bereich verschoben wird“ (S. 204). Aus musikpädagogischer Perspektive hat sich unter anderem Oberschmidt (2010) mit dem Einsatz von Metaphern im Musikunterricht auseinandergesetzt, ohne dabei Problemstellungen eines – in manchen Fällen vielleicht unbewussten – metaphorischen Sprachgebrauchs an Beispielen wie diesem zu ignorieren: „Auf dem Klavier liegen die hohen Töne ‚rechts‘, auf dem Cello ‚unten‘. Dem Laien mag metaphorisch erscheinen, was dem Fachmann als fixierter Terminus geläufig ist“ (S. 150).

Während Metaphern ein sprachliches Mittel darstellen beziehungsweise auf einer sprachlichen Ebene funktionieren, bezeichnen Analogien die „Ähnlichkeit oder Entsprechung zwischen zwei unterschiedlichen Dingen, Prozessen oder Sachverhalten“ (Kohl, 2007, S. 73). Aubussen et al. (2006) unterscheiden zwischen beiden Begriffen „in the sense that in metaphor, A is said to be B but in analogy, A is like B“ (S. 2). In einer Metapher wird also ausgesprochen, dass *A als B bezeichnet wird*, aber in einer Analogie, dass *A wie B ist*. Unklar bleibt hier die Unterscheidung zum Begriff *Vergleich*, da Kohl (2007) einen Vergleich mit derselben Aussage *A ist wie B* klassifiziert. Die Autorin führt aus: „Der Vergleich ist eine explizite Verbindung von zwei Vorstellungen oder Begriffen, die nicht identisch sind, aber (angeblich) in einem oder mehreren Aspekten eine

⁴⁶ Zum kontrovers geführten sprachwissenschaftlichen Diskurs um das Verhältnis zwischen *Analogie*, *Metapher* und *Vergleich* siehe zum Beispiel Coenen (2002) oder Kohl (2007). Die Darstellung innerhalb der vorliegenden Arbeit greift diesen nur rudimentär auf.

Ähnlichkeit aufweisen“ (Kohl, 2007, S. 73). Weiterhin klassifiziert sie den Vergleich gegenüber der Analogie als allgemeineren Begriff, der explizit keine Beweisführung hinsichtlich der Ähnlichkeiten nach sich zieht und weniger wissenschaftlichen Anspruch hat (Kohl, 2007).

Da das Hauptaugenmerk der vorliegenden Arbeit nicht nur auf sprachlichen Äußerungen beim unterrichtlichen Erklären im Musikunterricht, sondern auf der vermeintlichen Ähnlichkeit beziehungsweise (Un-)Gleichheit der gewählten Analogien liegt, wird der Begriff Analogie dem der Metapher vorgezogen. Auch hinsichtlich der Unterscheidung zwischen Analogie und Vergleich folgt die Darstellung dem Begriff Analogie, weil dadurch der Fokus mehr auf die didaktische Qualität der Ähnlichkeitsbeziehung gerichtet wird und nach Kohl (2007) „für didaktische und wissenschaftliche Kontexte angemessen“ (S. 76) erscheint. Der Begriff Analogie dient demnach als übergeordnete Klassifizierung, die jedoch auch Vergleiche und Metaphern innerhalb der Repräsentationsformen, die beim unterrichtlichen Erklären zur Anwendung kommen können, integriert. Analogien liegen damit auf einer anderen Ebene und quer zu Visualisierungen und Beispielen, da beiden Formen Analogien zugrunde liegen.

Shulman (1986) räumt dem Einsatz von Analogien („the most powerful analogies“, S. 9) beim unterrichtlichen Erklären als fachdidaktische Wissensfacette von Lehrkräften eine große Bedeutung ein und bezieht sich dabei auf empirische Studien, die den Einsatz hinsichtlich des Lernzuwachses von Schüler*innen als wirksam bestätigt haben. Harrison und Treagust (1993) untersuchten den Einsatz von Analogien im Physikunterricht der Sekundarstufe. Zentrale Ergebnisse der Studie zeigen, dass eine systematische Verwendung von Analogien zu mehr Verstehen bei den Schüler*innen führt (Harrison & Treagust, 1993). Dafür ist jedoch entscheidend, dass die Lehrkraft einerseits die richtige Interpretation der Analogie bei den Schüler*innen sicherstellt und andererseits vor allem die Aspekte der Analogie bespricht, die mit dem zu erklärenden Zielkonzept nicht übereinstimmen (Harrison & Treagust, 1993). Gerade das fehlende Eingehen auf die Elemente der Analogie, die ungleich des Zielkonzept sind, stellt auch Geelan (2013) als Ergebnis einer Videostudie fest: „(...) fewer paid careful and explicit attention to where the analogy broke down, or to the differences between the analogue and the target concept. It is possible that this sometimes introduced student misconceptions“ (S. 1759).

In der Musik spielen Analogien eine wichtige Rolle; vielen Begriffen, die in der Musik verwendet werden, liegen Analogien zugrunde. Beim Begriff Tonleiter beispielsweise handelt es sich um die Analogie einer Leiter, die man stufenweise herauf- und heruntersteigen kann, und einer diesbezüglich ‚auf-‘ beziehungsweise ‚absteigenden‘ Aneinanderreihung von Tönen innerhalb einer bestimmten Tonart. Auch Vortragsangaben wie *andante* (gehend) oder *con brio* (mit Feuer, lebhaft) liegen Analogien zugrunde, die den Charakter des musikalischen Vortrags in Worten auszudrücken versuchen, wie er von Komponist*innen intendiert ist. Auch bei der Vermittlung musiktheoretischer Inhalte werden Analogien zu Hilfe genommen. Beispielsweise dient der menschliche Puls als Analogie für den *Pulsschlag der Musik*, bei dem „aber von den Grundschlägen, vom Beat oder Metrum“ (Kemmelmeyer & Nykrin, 2004, S. 79) gesprochen wird. Die Ähnlichkeiten zwischen Pulsschlägen des Menschen und Grundschlägen der Musik bestehen in der

Regelmäßigkeit und den gleichen wiederkehrenden Abständen der Schläge. Weiterhin können beispielsweise bei der Vermittlung von Notenwerten oder Intervallen fachübergreifende Vergleiche zum Bruchrechnen im Fach Mathematik gezogen werden, etwa in der Darstellung des Abstandsverhältnisses zweier Töne in Form eines Bruches. Allerdings können diese Bezüge auch in umgekehrter Reihenfolge hergestellt werden, indem Notenwerte und Taktarten bei der Vermittlung von Bruchrechnen im Mathematikunterricht herangezogen werden (Krauss & Puffer, 2019). Entsprechende fachübergreifende Verweise können je nach Vorwissen der Schüler*innen auch zum Verstehen der Inhalte des eigenen Faches beitragen.

(4) Kombination von Repräsentationsformen

Zwischen den drei bisher genannten Formen der Repräsentation können Überschneidungen auftreten: Analogien können bei einer Erklärung nicht nur mündlich erwähnt, sondern dazu auch als Visualisierung abgebildet werden. Im vorherigen Beispiel aus dem Musikunterricht, der Darstellung verschiedener Tondauern sowie dem Verhältnis unterschiedlicher Notenwerte zueinander, können zwei ausgewählte Notenwerte als Beispiel dienen, deren Tondauer mit Hilfe eines Zeitstrahls visualisiert wird.

Generell empfiehlt Findeisen (2017) gerade bei der Vermittlung komplexer Sachverhalte, verschiedene Repräsentationsformen miteinander zu verbinden. Dabei können nicht nur „Beispiele und grafische Visualisierungen kombiniert“, sondern beispielsweise auch „unterschiedliche Visualisierungsformen verknüpft werden“ (Findeisen, 2017, S. 71; siehe auch Geelan, 2012; Schopf & Zwischenbrugger, 2015). Aus einer eher kognitionspsychologischen Betrachtungsweise heraus kann „die Aktivierung mehrerer Kanäle (z. B. auditiv und visuell) zu einer höheren Behaltensleistung führen und den Zugang für unterschiedliche Lerntypen ermöglichen“ (Findeisen, 2017, S. 69; siehe z. B. Brown & Hatton, 1982).

Darüber hinaus kommt aus musikpädagogischer Perspektive gerade der akustischen Repräsentationsform ein hoher Stellenwert zu (Frei et al., 2022), da das Musikhören beziehungsweise eher das Musikwahrnehmen – der Begriff hat sich im Laufe der Zeit als passender erwiesen (Laufer, 2018; siehe auch Niessen et al., 2008) – ein zentrales Element von Musikunterricht sein soll (siehe dazu z. B. Jank, 2013; Nimczik, 2013; Venus, 1969). Es ist daher anzunehmen, dass eine akustische, vor allem aber die Kombination aus akustischer und optischer Repräsentationsform bei der Vermittlung musiktheoretischer Inhalte sinnvoll ist, um diese Schüler*innen verständlich zu machen (Frei et al., 2022). Dies wird in Kapitel 2.4 ausführlich erläutert.

2.4 Erklären im Musikunterricht

Das vorliegende Kapitel richtet den Blick auf Erklären im Musikunterricht und fokussiert dabei Unterrichtsinhalte der Sekundarstufe I. Aufgrund der bisher weitgehend fachübergreifenden Herleitung von Qualitätsmerkmalen unterrichtlichen Erklärens ist die zentrale Frage, wie die bisherigen Ausführungen für das Erklären im Musikunterricht fruchtbar

gemacht werden können. Dafür werden zunächst solche Eigenschaften des Fachs Musik entfaltet, an denen sich eine Adaption fachübergreifender Aspekte unterrichtlichen Erklärens orientieren muss. Innerhalb dieses fachspezifischen Rahmens erfolgt zunächst die Unterscheidung von Erklären und Instruieren als Formen des unterrichtlichen Handelns von Musiklehrkräften. Auf Basis dessen wird dann ein Modell guten Erklärens beziehungsweise Instruierens im schulischen Musikunterricht entfaltet, das fachübergreifende Merkmale integriert und einen Schwerpunkt auf den Einsatz musikbezogener Repräsentationen legt. Die Modellierung orientiert sich dabei an der Vermittlung musiktheoretischer Inhalte, die als besonders *erklärbedürftig* erachtet werden und daher als Gegenstand für die Konzeptualisierung unterrichtlichen Erklärens im Fach Musik geeignet erscheinen.

2.4.1 Eigenschaften von Musikunterricht

Qualitätsmerkmale unterrichtlichen Erklärens einer Lehrkraft wurden in den bisherigen Ausführungen aus Mangel an musikpädagogischen Vorarbeiten vor allem aus einer fachübergreifenden Perspektive heraus entfaltet (siehe Kap. 2.3). Um diese für den Musikunterricht fruchtbar zu machen, sind zunächst Eigenschaften von Musikunterricht herauszuarbeiten, die sich von anderen Schulfächern unterscheiden und sich auf gutes Erklären im Musikunterricht beziehungsweise das Unterrichtshandeln von Musiklehrkräften auswirken können.

Einen ersten Anhaltspunkt bietet dabei die Forschung über domänenspezifische Kompetenzen von Musiklehrkräften. Nach Puffer (2021) stellen insbesondere drei fachliche Spezifika von Musikunterricht Herausforderungen für das Unterrichtshandeln von Lehrkräften dar: Sie ergeben sich erstens aus der „konstituierenden Bedeutung ästhetisch-musikalischer Praxis“, zweitens liegen „heterogene fachspezifische Lernvoraussetzungen von Schüler*innen“ vor; drittens beschreibt die Autorin Musikunterricht als gering strukturierte Domäne „mit einem unscharf umrissenen Unterrichtsgegenstand“ (S. 3), für den eine Vielzahl an Konzepten und Modellen vorliegt. Dabei lassen insbesondere die ersten beiden genannten Aspekte Annahmen über die Bedeutung des Erklärens im Musikunterricht zu.

Die Heterogenität in den Lernvoraussetzungen der Schüler*innen (Jank, 2013; Linn, 2017) lässt sich im Fach Musik unter anderem auf vielfältige außerschulische Bildungsangebote in Form von Gesangs- und Instrumentalunterricht sowie einem großen Angebot an Vokal- und Instrumentalensembles zurückführen (zusammenfassend siehe Ardila-Mantilla, 2018). Es kann daher vermutet werden, dass unterrichtliches Erklären im Fach Musik gerade zu einem differenzierten Angebot bei der Unterstützung aller Lernenden beitragen kann (siehe dazu Praetorius, Herrmann et al., 2020).

Der hohe Stellenwert ästhetisch-musikalischer Praxis im Musikunterricht setzt einen handelnden Umgang mit Musik voraus, der als Grundlage musikalischen Lernens angesehen wird (Fuchs, 2015; Jank, 2013) und bis heute als Qualitätsmerkmal von Musikunterricht gilt (Krause-Benz, 2016). Nach Stadler-Elmer (2000) sind drei Bereiche musikalischen Handelns grundlegend, um musikalische Fähigkeiten zu entfalten: Sich zur Musik

bewegen, Musik Hören und Klänge erzeugen. Ein derartiges Handeln kann – ganz in der Tradition Aeblis (1983) – als Ausgangspunkt für einen offen gedachten Prozess musikalischen Lernens gesehen werden (siehe Jank, 2013). Kaiser (2010) spricht vom Leitziel einer Verständigen Musikpraxis und reklamiert für das Handeln einer Musiklehrkraft eine konstruktive Absicht, die es Schüler*innen ermöglicht, „eine vorhandene Usuelle in eine Verständige Musikpraxis überführen zu können“ (S. 66). Ähnlich argumentiert Rolle, der von einer Musiklehrkraft fordert, sie müsse „vielfältige Räume für musikalisches Handeln eröffnen, in denen ästhetische Erfahrungen möglich sind, angeregt und unterstützt werden“ (Rolle, 1999, S. 5).

Für das Erklären im Musikunterricht bedeutet das Prinzip der Handlungsorientierung daher eine vermutlich weniger prominente Rolle im Gegensatz zu stärker kognitiv orientierten Fächern wie Mathematik. Dies wird beispielsweise auch deutlich, wenn Fuchs (2015) „theoretische Erklärungen“ im Musikunterricht zwar als „sinnvoll und notwendig“ bezeichnet, dies aber unbedingt voraussetze, dass Schüler*innen bereits „über ausreichende musikalische Erfahrungen verfügen“ (S. 102). Davon abgeleitet wird die eindeutige Prämisse: „Handlung vor Wissen – Klang vor Zeichen“ (Fuchs, 2015, S. 102). Auch Gebauer (2013) beschreibt als typische Situation von Musikunterricht einen musikpraktischen Stundenbeginn, der vor dem Lernen von Begriffen ein reines Wahrnehmen des musikalischen Phänomens ermöglicht.

Ein weiteres konstituierendes Moment für musikpädagogische Praxis⁴⁷ stellen nach Godau (2018) die Dinge⁴⁸ dar, die im Musikunterricht zum Einsatz kommen. Kranefeld (2021) betont die – auch im musikpädagogischen Forschungsdiskurs – zunehmende Relevanz der „materialen Dimension des Unterrichts“ (S. 229) als mögliches Merkmal von Unterrichtsqualität im Fach Musik. Sie verweist auf das Bild des didaktischen Tetraeders (Hußmann et al., 2018), der die Idee des didaktischen Dreiecks (Lehrende, Schüler*innen und Lerngegenstand; siehe dazu Reusser, 2003) aufgreift und für den Musikunterricht um die vierte Komponente der Artefakte erweitert. Kranefeld et al. (2019) arbeiten die Bedeutung der Verwendung von Dingen und Artefakten für Prozesse des Musik-Erfindens von Schüler*innen heraus. In Bezug auf das Unterrichtshandeln einer Musiklehrkraft sind insbesondere solche Dinge relevant, die beim Erklären zum Einsatz kommen können (z. B. eine Tafel beziehungsweise ein Whiteboard mit Notenlinien oder Instrumente zur Darstellung von Klängen).⁴⁹

Zusammenfassend sind für das Erklären im Musikunterricht folgende Aspekte wichtig: Erstens stellt der handelnde Umgang mit Musik ein entscheidendes Merkmal von

47 Gemeint ist hier das Unterrichten als generelle Praxis und nicht, wie vielleicht zu vermuten, die rein musikalische Praxis in Form gemeinsamen Singens oder Musizierens als Element von Musikunterricht.

48 Dieser Terminus wird im Kontext der Materialität von (Musik-)Unterricht nicht nur im musikpädagogischen, sondern auch im erziehungswissenschaftlichen Diskurs verwendet (z. B. Jörissen, 2015).

49 In der GeiGe-Studie zeigte sich bezüglich der Inszenierungstechniken von Instrumenten durch Musiklehrkräfte, dass es eine große Rolle spielen kann, inwieweit die Lehrkraft das Instrument selbst beherrscht, und davon gegebenenfalls abhängt, ob Instrumente „eher als physikalische Klangerzeuger denn als musikalische Klang- und Ausdrucksträger inszeniert“ (Kranefeld, Heberle et al., 2015, S. 159) werden.

Musikunterricht dar, um Musik als klingendes Phänomen zu begreifen. Gutes Erklären im Musikunterricht sollte demnach mit klingender Musik einhergehen. Zweitens wird auch den Materialien, die im Musikunterricht zum Einsatz kommen, eine hohe Bedeutung beigemessen. Gutes Erklären im Fach Musik sollte daher die *Dinge* reflektiert einsetzen. Musikunterricht zeichnet sich drittens aufgrund außerschulischer Angebote durch große Heterogenität zwischen den Lernenden aus. Gutes Erklären im Musikunterricht sollte demnach die Heterogenität zwischen den Schüler*innen berücksichtigen. Die drei genannten Aspekte können sich darüber hinaus wechselseitig beeinflussen, etwa hinsichtlich der Frage, ob für alle Schüler*innen beim Erklären gleichermaßen wichtig ist, dass Musik erklingt, oder ob dabei unterschiedliche musikalische Erfahrungen eine Rolle spielen. Weiterhin erheben sie keinen Anspruch auf Vollständigkeit, dienen aber für die folgende Darstellung als grundlegende fachspezifische Rahmung.

2.4.2 Erklären und Instruieren

Die Frage nach der Unterscheidung der beiden Handlungen Erklären und Instruieren⁵⁰ im Kontext von Musikunterricht wurde bereits im Rahmen der domänenspezifischen Konzeptualisierung fachdidaktischen Wissens von Musiklehrkräften gestellt (siehe Kap. 2.2.2). Die in der COACTIV-Studie mit den Begriffen Erklären, Darstellen, Repräsentieren oder Verständlichmachen versehene fachdidaktische Wissensfacette weisen Puffer und Hofmann (2017) in FALKO-M als „Wissen um musikbezogene Instruktionsstrategien“ aus (S. 254). Im Gegensatz zu den weiteren in FALKO beteiligten Fächern verzichten sie dabei explizit auf den Terminus Erklären (siehe Krauss et al., 2017), verwenden aber den Begriff „Instruktion“ als umfassenden Sammelbegriff (siehe Kap. 2.2.2). Unter diesem Begriff werden Erklärhandlungen zwar nicht ausgeschlossen, allerdings erscheint der Begriff Erklären auch nicht hinreichend, um das Unterrichtshandeln von Musiklehrkräften in Vermittlungssituationen umfassend zu beschreiben. So wird beim Blick auf Musizierprozesse deutlich, dass das Anleiten von Gestaltungsprozessen weniger einer erklärenden Lehrkraft als vielmehr einer instruierenden bedarf (siehe z. B. Fuchs, 2015).

Im Gegensatz dazu lenken Aspekte wie der Einsatz passender Illustrationen oder die Demonstration von Lerninhalten mit anhörlichen Beispielen (Puffer & Hofmann, 2017) den Blick auch auf kognitive Vermittlungskontexte, in denen erklärt wird. Für die Modellierung von Erklären im Fach Musik kann die Konzeptualisierung in FALKO-M somit als fachspezifischer Anhaltspunkt dienen, der jedoch eine Differenzierung zwischen Erklären und Instruieren erforderlich macht. Die Unterscheidung beider Handlungen orientiert sich am Diskurs der Linguistik, die zwischen beiden Begriffen anhand des dabei verfolgten Zwecks trennt (z. B. Hohenstein, 2009). Die folgende Darstellung greift diesen

50 Der in Kapitel 2.1 erwähnte Begriff des instruktionalen Erklärens ist vor dem Hintergrund der Unterscheidung beider Sprachhandlungen zu spezifizieren. Er meint im hier diskutierten Kontext keine Mischform von Erklären und Instruieren, sondern ist aus dem Englischen abgeleitet. Er wird dort als *instructional explanation* (Leinhardt, 2001) gebraucht und bezieht sich auf ein Erklären in Lehr- und Lernkontexten.

Diskurs auf und leitet zur Modellierung guten Erklärens und Instruierens, die bereits an anderer Stelle vorgeschlagen wurde (Frei et al., 2022).

Instruieren

Nach Becker-Mrotzek (2004) sind Instruktionen „komplexe Handlungen“, die „der Vermittlung von praktischen Fertigkeiten dienen“ (S. 131) und anderen Personen Handlungskonzepte vermitteln – nach Rehbein (1977) sind damit Fähigkeiten zur Durchführung von Handlungen gemeint. Hohenstein (2009) sieht in Instruktionen den Zweck, andere zum Selber-Handeln zu befähigen. Grundsätzlich kann zwischen den beiden Formen mündliches Instruieren und schriftliches Anleiten⁵¹ unterschieden werden (Becker-Mrotzek, 2004). Für das Unterrichtshandeln von Musiklehrkräften sind diesbezüglich vor allem mündliche Instruktionen beim gemeinsamen Singen, Musizieren und sich zur Musik bewegen relevant (z. B. Stibi, 2019).

Allerdings sind innerhalb musikpraktischer Vermittlungskontexte auch Situationen denkbar, in denen eine Musiklehrkraft nicht nur durch ihre Stimme (z. B. durch Einbeziehungweise Vorzählen), sondern auch durch Gesten (z. B. durch Dirigieren; siehe z. B. Leeb, 2021; Rucha, 2016; Schäfer, 2021) oder durch den Einsatz von Instrumenten (z. B. schulpraktisches Klavier- und Gitarrenspiel; siehe z. B. Puffer, 2021) das gemeinsame Singen der Klasse in Gang setzt, anleitet und koordiniert⁵². Instruktionshandlungen einer Musiklehrkraft sind daher nicht ausschließlich mündlich zu verstehen, sondern können sich in verschiedenen Instruktionsmodi vollziehen. Kurze Erklärungen können in diesen Kontexten zwar vorkommen, spielen jedoch eine eher untergeordnete Rolle (z. B. Eberhard, 2010).

Diese Möglichkeit bestätigt Neumeister (2011) auch aus linguistischer Perspektive, nach der es innerhalb von Instruktionen zu „unterschiedlichen Sprechhandlungen, u.a. auch zum Erklären, kommen kann“ (S. 61). Für das Unterrichtshandeln einer Musiklehrkraft wäre dabei in Anlehnung an Hohenstein (2009) relevant, die zu erlernende Handlung in einzelne Schritte zu zerlegen, um Schüler*innen zur eigenen Handlungsdurchführung zu befähigen. Ein Situationsbeispiel aus dem Musikunterricht könnte sein, dass eine Musiklehrkraft beim Erlernen einer Melodie kurze Elemente vorsingt und im Call-Call-Prinzip immer dann weitere Teile der Melodie hinzunimmt, wenn die Schüler*innen einen Teil richtig nachgesungen haben. Damit Schüler*innen eine solche musikalische Handlungskompetenz entwickeln können, ist ein implizites, prozedurales Handlungswissen notwendig (Gruhn, 2003b; siehe auch Hohenstein, 2009).

Erklären

Erklären hat nicht nur das Ziel, neues Wissen aufzubauen und innere Zusammenhänge von Wirkungs- oder Funktionsweisen zu verdeutlichen (Becker-Mrotzek, 2004), sondern

51 Siehe hierfür zum Beispiel Aeschbacher (2009) oder Wellenreuther (2013).

52 Zum Begriff *Koordination* in Bezug auf Musikpädagogik siehe z. B. Spychiger (2019).

auch neue noch unverbundene Informationen in vorhandenes Wissen integrieren beziehungsweise in einen größeren Gesamtzusammenhang einordnen zu können (Hohenstein, 2006 und 2009). Beim Erklären wird demnach der Zweck verfolgt, das Wissen der Adressat*innen auf- beziehungsweise auszubauen (Hohenstein, 2009). Auf didaktische Handlungskonzepte angewandt, definiert Kiel das Verstehen auf Seiten der Schüler*innen als „die zentrale und notwendige Funktion des Erklärens“ (S. 83). Analog zu anderen Fächern wie Deutsch oder Mathematik, die sich dieser zentralen Funktion anschließen (z. B. Neumeister, 2011; Wörn, 2014), bedeutet das auch für das Erklären im Musikunterricht, dass insbesondere kognitive Ziele verfolgt werden.

Diese Setzung wird für die vorliegende Unterscheidung vorgenommen und erscheint aus den folgenden Gründen notwendig: Im musikpädagogischen Schrifttum findet sich ein umfassender Diskurs zum *Musik verstehen* (z. B. Flämig, 1998; Rauhe et al., 1975), den Kraemer (2007) anhand verschiedener Bedeutungszuschreibungen zusammenfasst. So ist von *Musik verstehen* beispielsweise auch die Rede, wenn die Botschaft einer Komposition erkannt und entschlüsselt oder ein Werk von Musiker*innen adäquat erfasst und interpretiert werde (Kraemer, 2007). In beiden Fällen vollzieht sich eine solche Analyse jedoch unter Gültigkeit einer individuellen Bedeutungserschließung der Rezipient*innen, weshalb ein objektiver Begriff von Verstehen hier unpassend erscheint und einer kognitiven Ausrichtung nicht entspricht.

Heß (2013) überträgt ein solches Verständnis von *Musik verstehen* auf Situationen im Musikunterricht. Demnach könnten Aussagen von Schüler*innen wie *Ich verstehe diese Musik nicht* beim Hören klassischer Musik darauf zurückzuführen sein, dass sie einen „spontanen und lustvollen Umgang“, wie sie ihn mit dem Hören populärer Musik verbinden, als unzureichend empfinden und „glauben, ihr damit nicht gerecht zu werden. (...) Verstehen scheint in diesem Denken auf Analysieren und Interpretieren angewiesen zu sein“ (S. 205). Daher entspricht Verstehen im Sinne einer subjektiven Interpretation von Musik nicht den Verstehensprozessen, wie sie als Ziel unterrichtlichen Erklärens in Kapitel 2.3.1 definiert wurden. Auch bestünde dann kein komplementäres Verhältnis zwischen Erklären und Verstehen mehr, wie es von Kiel (1999) skizziert wurde.

Vor dem Hintergrund der heterogenen Bedeutungszuschreibungen im musikpädagogischen Diskurs fokussiert die vorliegende Arbeit ein Verständnis von unterrichtlichem Erklären, das zu einem *Musik verstehen* im Sinne eines Wissensaufbaus bei Schüler*innen führt – andere nicht kognitive Aspekte sind hier nicht einbezogen. In der Gegenüberstellung von Kraemer (2007) ist ein solches Verstehen im Kontext des „Erfassens struktureller Zusammenhänge“ (S. 101) zu verorten. Im Zuge dessen erachtet der Autor unter anderem musiktheoretische Kenntnisse als notwendige Voraussetzung, um beispielsweise formale Verläufe oder harmonische Folgen erfassen sowie Instrumentierungen, Gattungen oder musikalische Zitate erkennen zu können. Beim Erklären verfolgt eine Musiklehrkraft demnach immer auch das Ziel, Schüler*innen solche Kenntnisse zu vermitteln. Dieser Ansatz ist mit dem kognitionspsychologisch orientierten Begriff von Verstehen gut vereinbar, der das Zielkriterium unterrichtlichen Erklärens darstellt (siehe dazu die *Definition unterrichtlichen Erklärens einer Lehrkraft*, Kap. 2.2.2).

Erklären(-Wie) und Instruieren

In den bisherigen Ausführungen wurde Erklären und Instruieren auf theoretischer Ebene getrennt voneinander dargestellt. Vor dem Hintergrund unterschiedlicher Typologien, die zwischen Erklären-Was, Erklären-Wie und Erklären-Warum unterscheiden (Klein, 2009; siehe Kap. 2.3.1), bedarf es einer weiteren Klärung zwischen Instruieren und Erklären-Wie. Beide Handlungen haben gemein, insbesondere prozedurales Wissen bei den Adressat*innen zu entwickeln (siehe Hohenstein, 2009). Das könnte auch einer der Gründe sein, weshalb selbst in älteren linguistischen Arbeiten gerade das Erklären-Wie mit anderen sprachlichen Handlungen wie Beschreiben oder Instruieren gleichgesetzt wird (siehe Rathausky, 2010; siehe auch Kap. 2.1). Bei genauerer Betrachtung sind jedoch auch hier Unterschiede festzustellen, die das Unterrichtshandeln von Musiklehrkräften weiter ausdifferenzieren können. Nach Gruhn (2003b) ist es möglich, Schüler*innen explizites Wissen über die Funktion der Stimme zu vermitteln, ohne bei ihnen dadurch implizites Handlungswissen zu entwickeln – *wie* die Stimme funktioniert, kann eine Lehrkraft erklären. Allerdings würde man von Instruieren sprechen, wenn eine Lehrkraft verschiedene Funktionsweisen der Stimme darstellt, indem sie den Schülern eine Melodie vorsingt und diese so zum Nachsingen befähigt. Die Lehrkraft zeigt dabei Funktionszusammenhänge der Stimme nicht explizit auf, sondern vermittelt sie implizit.

Weniger klar wird das Bild bei der Vermittlung musiktheoretischer Inhalte. Beim Blick in Methodensammlungen und musikdidaktische Publikationen⁵³ fällt auf, dass der Begriff Erklären kaum zu finden ist – bei diesen seltenen Fällen kommt es dann auch vor, dass diese Aufgabe Schüler*innen zugeschrieben wird, die sich etwas selbst erklären sollen (Eberhard, 2016). Häufiger begegnet man in Zusammenhang mit Empfehlungen für das Unterrichtshandeln dem Begriff *Erläutern*, der bei genauerer Wortbetrachtung jedoch als Synonym für *Erklären* zu lesen ist (z. B. Buschendorff, 2010; Niessen, 2010). Darüber hinaus finden sich an manchen Stellen methodische Hinweise, denen jedoch eher ein *Instruieren* anstatt eines Erklärens(-*Wie*) zugrunde liegt – mit der aus der Wortbedeutung heraus theoretisch abzuleitenden Folge, dass zwar Handlungsvollzüge auf Seiten der Instruierten möglich sind, die jeweiligen Wissenszusammenhänge aber ohne ein Erklären-Wie vermutlich nicht deutlich werden.

Beispielsweise gibt Nowak (1999) bei der Einführung in das Bilden von Akkordumkehrungen Prozessschritte ohne weitere inhaltliche Erklärung der Zusammenhänge zwischen Umkehrung und dem jeweils am tiefsten klingenden Ton des Akkordes Folgendes vor: „Oktaviert man den untersten Ton eines Dreiklangs, erhält man keinen neuen Akkord, sondern lediglich seine Umkehrung.“ (S. 31). Beschrieben wird ein Handlungsprozess, der Schüler*innen zum Nachvollzug befähigen kann. Wenn allerdings notwendige Zusammenhänge nicht explizit erklärt werden, darf eine inhaltliche Durchdringung und ein nachhaltiger Wissensaufbau bezweifelt werden. Um Wissen der Schüler*innen aufzubauen, wäre beispielsweise die Erklärung notwendig, dass bei der Bildung der ersten Umkehrung der unterste Ton aus dem Grund eine Oktave höher geschrieben wird, weil

53 Beide Begriffe *Methodensammlung* und *musikdidaktische Publikationen* orientieren sich an der Zusammenfassung von Jank (2018, S. 358–359).

dadurch der Terzton zum am tiefsten klingenden Ton wird – dies stellt ein klangliches wie funktionales Charakteristikum der ersten Umkehrung dar.

Ohne diese Informationen erwerben Schüler*innen durch instruierendes Handeln der Lehrkraft vermutlich vor allem prozedurales Wissen, das jedoch ohne weitere Information noch nicht in deklaratives Wissen – wie in diesem Fall Kenntnisse über Charakteristika von Akkordumkehrungen – überführbar ist. Im Gegensatz dazu wird angenommen, dass instruierendes Handeln für das Ziel ausreichend ist, Schüler*innen zum Nachsingen einer einfachen und kurzen Melodie zu befähigen. Sie benötigen dafür weder explizites und systematisches Wissen über Struktur oder Rhythmus dieser Melodie noch weitere andere Hintergrundinformationen, um die Melodie nachsingen zu können. Hinsichtlich der Klassifikation beider Handlungen kann man mit Neumeister und Vogt (2009) sagen, dass das Erklären-Wie über das Instruieren hinausgehe, indem es nicht nur Prozessabläufe aufzeigt, sondern sich diesbezüglich auch auf Prozesshintergründe beziehungsweise konkrete Wissens Elemente bezieht.

Zusammenfassend lässt sich hinsichtlich des *Erklärens* und *Instruierens* im Kontext von Musikunterricht konstatieren, dass beide Handlungen im Repertoire des Unterrichtshandelns von Musiklehrkräften eine wichtige Rolle spielen. Auf theoretischer Ebene wird anhand des dabei verfolgten Zwecks unterschieden, ob von Erklären oder Instruieren gesprochen werden muss. Während sich beide Handlungen also theoretisch voneinander trennen lassen, erfolgt die praktische Umsetzung in Mischformen (siehe für eine solche Situation z. B. Brandstätter, 2010). Dabei kann genauso von Erklärprozessen gesprochen werden, in denen Instruktionen vorkommen, als auch von Instruktionsprozessen, in denen kurze Erklärungen Platz haben. Mit Blick auf die Erklärkompetenz einer Lehrkraft kann die Konzeptualisierung musikbezogener Instruktionsstrategien von Puffer und Hofmann (2017) daher sprachlich geschärft werden, indem der Terminus *musikbezogene Strategien im Erklären und Instruieren* verwendet wird (Frei et al., i. Vorb.). Dadurch werden einerseits Spezifika von Musikunterricht berücksichtigt, andererseits nähert sich die Bezeichnung auf begrifflicher Ebene wieder der ursprünglichen und als Grundlage dienenden Taxonomie Shulmans (1986) an.

Zwischenfazit

Als Zwischenfazit bedeuten die bisherigen Ausführungen für die Konzeptualisierung guten Erklärens im Musikunterricht zweierlei: Einerseits scheint es sinnvoll, Erklären als Unterrichtshandeln einer Lehrkraft in enger Verknüpfung zu Instruieren zu modellieren (siehe Frei et al., 2022); andererseits verfolgen beide Handlungen unterschiedliche Ziele und können anhand derer klar voneinander getrennt werden. Wenn also Erklären – in Anlehnung an Paulis (2015) Zitat zu Beginn – auch eine Kerntätigkeit für Musiklehrkräfte darstellt, dann sind damit Handlungen gemeint, die den primären Zweck des Wissensaufbaus der Schüler*innen verfolgen. Die Modellierung guten Erklärens und Instruierens integriert dabei sowohl die bisher vorwiegend fachübergreifend hergeleiteten und auf den Musikunterricht gewendeten Qualitätsmerkmale (siehe Kap. 2.3.2) als auch die fachspezifische Ausrichtung auf Basis der Strategien, die aus FALKO-M abgeleitet wurden

(Puffer & Hofmann, 2017). Der Fokus liegt dabei insbesondere auf der musikspezifischen Auslegung des Einsatzes von Repräsentationen.

Modell guten Erklärens bzw. guten Instruierens im Musikunterricht

Aufgrund des Desiderats zum unterrichtlichen Erklären in der musikpädagogischen Forschung wurden sowohl allgemeine Merkmale guten Erklärens aus dem Diskurs um Unterrichtsqualität adaptiert als auch Aspekte aus anderen Domänen entfaltet und für das unterrichtliche Erklären im Musikunterricht fruchtbar gemacht. Dabei wird angenommen, dass alle genannten Qualitätsmerkmale (Adressatenorientierung, Strukturiertheit, Sprachliche Verständlichkeit, Sprech- und Körperausdruck sowie der Einsatz von Repräsentationen) auch für das Erklären im Musikunterricht gelten (Frei et al., 2022).

Allerdings zeigte sich bei der fachübergreifenden Spezifizierung, dass wichtige Formen externer Repräsentation fehlen, die für den Musikunterricht notwendig sind. Musik als klingendes Phänomen rückt demnach insbesondere *akustische Repräsentationen* ins Zentrum. Während beispielsweise bei Vokabelerklärungen im Englischunterricht der klanglichen Realisation neben weiteren nonverbalen Faktoren, wie unterstützender Gestik, eine durchaus relevante Rolle beigemessen wird (Lazaraton, 2004), stellt sie für Erklärungen im Musikunterricht ein zentrales Element dar: Klangliche Realisationen als bedeutsame Form musikalischer beziehungsweise musikunterrichtlicher Kommunikation (z. B. Hellberg, 2018) erfordern die Berücksichtigung akustischer Repräsentationen beim Erklären und weisen ihnen eine zentrale Rolle zu.

Puffer und Hofmann (2017) benennen mit Blick auf das dabei notwendige Professionswissen von Musiklehrkräften Formen musikpraktischen Vormachens wie Vorsingen, Vorspielen eines Instrumentes oder auch Vorklatschen. Sie führen neben dem Wissen um anschauliche Illustrationen auch das Wissen um anhörliche Beispiele an (Puffer & Hofmann, 2017). Diese Klang- beziehungsweise Hörbeispiele können musikalische Phänomene klanglich repräsentieren und sollten demnach beim Erklären eingesetzt werden. Als Beispiel soll hier der Aufbau einer Dur-Tonleiter dienen: Vermittelt werden kann diese, indem die Tonleiter notiert und mithilfe des musikalischen Zeichensystems optisch repräsentiert wird. Zur weiteren Veranschaulichung können auch die im Wort steckenden Analogien einer Leiter beziehungsweise einer Treppe herangezogen werden (z. B. Deterbeck & Schmidt-Oberländer, 2019, S. 72). Um das Auf- und Absteigen von leitereigenen Tönen jedoch auch akustisch zu fassen, können diese durch ein Instrument von der Lehrkraft oder mit Hilfe von Boomwhackers durch Schüler*innen klanglich repräsentiert werden.

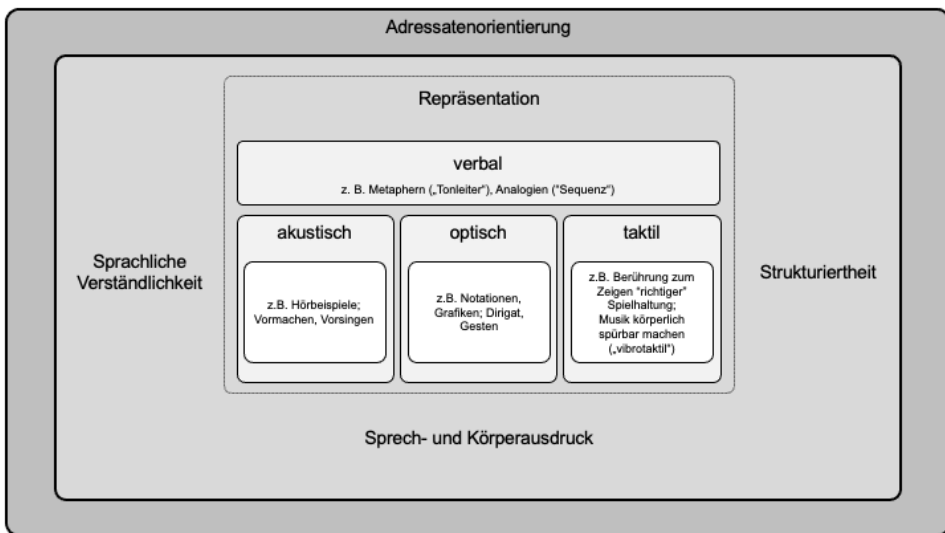
Vor dem Hintergrund sensomotorischer Prozesse beim Singen und Musizieren können auch *taktile* Repräsentationen zum Einsatz kommen. Kinästhetische oder haptische Wahrnehmungsprozesse der Schüler*innen können durch gezielte Berührungen in Gang gesetzt werden (Schröter, 2021; Steffen-Wittek, 2021): Beispielsweise können die Schüler*innen beim Singen verschiedene Resonanzräume im Nasen- und Rachenraum berühren und so Klang als Form von Schwingung auf einer anderen Ebene als einer akustischen Repräsentation wahrnehmen. Auch können Lehrpersonen durch gezielte Berührungen bei

singenden oder musizierenden Schüler*innen Wahrnehmungsprozesse hinsichtlich Körperhaltung beziehungsweise Einbindung des gesamten (Klang-)Körpers adressieren (Frei et al., 2022).⁵⁴

Unklar ist in diesem Zusammenhang jedoch, welche Bedeutung die einzelnen Repräsentationen für das Erklären im Fach Musik haben. Orientiert man sich an der Empfehlung von Fuchs (2015), die als ein Prinzip guten Musikunterrichts die Prämisse „Handlung vor Wissen – Klang vor Zeichen“ (S. 102) anführt, kann für den Einsatz von Repräsentationen beim Erklären im Musikunterricht Folgendes abgeleitet werden: Es sollte erst erklärt werden, nachdem Musik erklingen ist; zudem sollte der akustischen Repräsentation beziehungsweise dem eigenen Musizieren ausreichend Zeit eingeräumt werden (Fuchs, 2015). Allerdings liegen empirische Erkenntnisse weder zur Wirksamkeit bestimmter Reihenfolgen noch Kombinationen von Repräsentationsformen vor. Während Fragen zur Wirksamkeit in dieser Arbeit nicht beantwortet werden können, werden jedoch erste Anhaltspunkte darüber gewonnen, welche Bedeutung die einzelnen Formen musikbezogener Repräsentation für verschiedene Statusgruppen haben.

In das nachfolgende Modell fließen sowohl die Merkmale guten Erklärens, die fachübergreifende Gültigkeit haben (siehe Kap. 2.3.2) als auch der Einsatz musikbezogener Repräsentationen ein, die in unterschiedlichen Formen realisiert werden können (zum Aufbau des Modells siehe Frei et al., 2022).

Abbildung 2.2 Modell guten Erklärens bzw. Instruierens im Musikunterricht (Frei et al., 2022, S. 205)



54 Diese Komponente aus dem Modell guten Erklärens und Instruierens im Musikunterricht wird in der vorliegenden Studie nicht untersucht, da insbesondere Aspekte im forschungsmethodischen Vorgehen eine valide Erhebung nicht möglich machen (siehe Kap. 4).

Adressatenorientierung ist als grundlegendes Leitprinzip (siehe Kap. 2.3.2) und *Hintergrundfolie* einer Erklärung hinterlegt. Es stellt ein tiefenstrukturelles Merkmal unterrichtlichen Erklärens dar. Sprachliche Verständlichkeit, Strukturiertheit, Sprech- und Körperausdruck sowie der Einsatz von Repräsentationen bilden zusammen eine weitere Ebene. Dabei ist jedoch zum einen zu berücksichtigen, dass Strukturiertheit und sprachliche Verständlichkeit eher tiefenstrukturelle Merkmale darstellen, während Sprech- und Körperausdruck sowie der Einsatz von Repräsentationen eher auf der Oberflächenstruktur unterrichtlichen Erklärens zu verorten sind. Zum anderen wird zwischen verbalen und nonverbalen (optisch, akustisch und taktil) Repräsentationen unterschieden.

Erklären und Instruieren im Lichte musikdidaktischer Handlungsfelder

Diese theoretische Modellierung guten Erklärens beziehungsweise Instruierens im Musikunterricht soll im Folgenden durch Beispiele aus dem Musikunterricht veranschaulicht und erläutert werden. Ziel ist dabei nicht die Differenzierung zwischen Erklären und Instruieren anhand verschiedener musikdidaktischer Handlungsfelder (Frei et al., 2022; siehe Dartsch et al., 2018, Kap. 4.6.2). Vielmehr dienen diese für eine übersichtliche Darstellung möglichst vielfältiger Beispiele aus dem Musikunterricht, die gleichzeitig verschiedene Umgangsweisen mit Musik (Venus, 1969) berücksichtigen. Auch kann dadurch gewährleistet werden, eine möglichst große Bandbreite von Musikunterricht abzubilden, in dem unterrichtliches Erklären eine Rolle spielen kann.

Um den Fokus auf das Unterrichtshandeln von Musiklehrkräften zu legen, richtet sich der Blick dabei auf methodische Aspekte des Unterrichtens (z. B. Fiedler & Handschick, 2014; Schulten, 1980). Da nicht nur der musikpädagogische, sondern auch der bildungswissenschaftliche Diskurs in vielfältiger Weise auf den Begriff Methode zurückgreift und dessen Bedeutung sich nach Jank (2018) „nur aus dem konkreten Gebrauch im jeweiligen Kontext erschließen“ (S. 355) lässt, wird dieser zunächst knapp erläutert. Unter Rekurs auf Meyer (2004) unterscheidet Helmke (2017) beim Begriff Methode zwischen drei Ebenen: Während er methodische Großformen wie Frei- oder Projektarbeit auf einer Makroebene und Sozialformen (Plenumsunterricht, Gruppenunterricht, Einzelarbeit) oder Handlungsmuster (Tafelarbeit, Vortrag) als Dimensionen methodischen Handelns auf einer Mesoebene positioniert, ordnet er Inszenierungstechniken wie Zeigen, Modellieren, Vormachen oder Verfremden auf einer Mikroebene an (Helmke, 2017). Die musikdidaktische Literatur greift diese Einteilung auf und nutzt sie als strukturelle Grundlage (z. B. Biegholdt, 2013; Eberhard, 2016), wobei ein didaktischer Zusammenhang in solchen Publikationen beziehungsweise Methodensammlungen nach Jank (2018) nur ansatzweise erfasst wird. Nach dieser Definition ist das methodische Handeln einer Lehrkraft, die im Musikunterricht etwas erklärt, grundsätzlich auf der Mikroebene zu verorten. Da beim Erklären jedoch auch andere Inszenierungs- beziehungsweise Vermittlungstechniken wie Vormachen oder Verfremden in den Erklärprozess eingebaut werden können, stellt sich die Frage, ob ein solcher Prozess nicht auch als Handlungsmuster angesehen werden kann. Unabhängig dieser Ebenen wird in dieser Arbeit Erklären als Prozess verstanden, in dem auch andere Techniken zum Einsatz kommen können.

Grundlage der folgenden Darstellung sind die Ausführungen im *Handbuch Musikpädagogik* (Dartsch et al., 2018), das verschiedene Perspektiven zum Komplex Musiklernen einnimmt. Eine davon unterscheidet aus der Sicht des komplementären Themenfelds ‚Lehren‘ verschiedene musikdidaktische Handlungsfelder, „für die jeweils eigene didaktische Traditionen und Fachszenen existieren“, wobei in der Darstellung ein „Verlust von Trennschärfe zwischen den Handlungsfeldern in Kauf genommen“ (Dartsch & Platz, 2018, S. 158) wird. Allerdings spiegelt gerade ein Überlappen beziehungsweise Ineinandergreifen mehrerer Handlungsfelder die Komplexität und Vielfältigkeit von Musikunterricht wider. Da die verschiedenen Autorinnen und Autoren jedoch an manchen Stellen eine gesamt-musikpädagogische und nicht nur auf Musikunterricht an allgemeinbildenden Schulen ausgerichtete Perspektive einnehmen (siehe Dartsch et al., 2018), werden im Folgenden vor allem Aspekte fokussiert, die für schulischen Musikunterricht relevant sind. Fünf Lernbereiche werden definiert (siehe Dartsch et al., 2018):

- (1) Singen und instrumentales Musizieren
- (2) Tanzen und sich zur Musik bewegen
- (3) Improvisieren und Komponieren von Musik
- (4) Wahrnehmen von Musik
- (5) Nachdenken und Sprechen über Musik

Für eine übersichtliche Darstellung wird ein wiederkehrender Zweischritt angewendet: Zunächst erfolgt eine kurze Skizzierung des jeweiligen Handlungsfeldes, bevor daran anschließend Situationen entfaltet werden, in denen Erklären und Instruieren einer Lehrkraft im Musikunterricht zum Einsatz kommen kann.

(1) Singen und instrumentales Musizieren

Das vokale und instrumentale Musizieren bereits komponierter Musik stellt neben einer Vielzahl an außerschulischen musikalischen Praxen (z. B. Dartsch, 2018; Pezenburg, 2018) ein „besonders bedeutsames Moment“ (Puffer & Hofmann, 2017, S. 256) für die Gestaltung von Musikunterricht dar. Zwar gibt es hinsichtlich des Musizierens innerhalb einer Klassengemeinschaft – häufig mit dem Begriff Klassenmusizieren⁵⁵ bezeichnet – weit gefasste Definitionen, die unter gemeinsamer Musizierpraxis und neben Singen und Instrumentalspiel auch gemeinsame Bewegung und szenisches Spiel integrieren (z. B. Eberhard, 2016). Jedoch richtet sich der Fokus hier auf die beiden Handlungsfelder Singen und Instrumentalspiel – auch um Unklarheiten zu vermeiden (Papst-Krueger, 2013).

Nach Nimczik (2002) gilt Singen als „ältester Gegenstand“ (S. 4) des Schulfachs Musik. Diese Beobachtung bekräftigt Dyllick (2018) und begründet dies unter anderem anhand der bis in die 1920er Jahre gültigen Bezeichnung *Gesangunterricht* beziehungsweise in Form eines in heutiger Zeit stetig wachsenden Angebots an Chorklassen. Einen gerade für den Musikunterricht praktikablen wie pragmatischen Vorteil „bietet Singen durch die allgemeine Verfügbarkeit der Stimme gegenüber anderen, instrumentalen Formen des Musizierens“ und ermöglicht einer Musiklehrkraft dadurch, „prinzipiell alle

55 Zur unterschiedlichen Verwendung des Begriffs siehe zusammenfassend beispielsweise Papst-Krueger (2013) und Heß (2018a).

Schüler*innen nahezu voraussetzungslos ins Musizieren miteinbeziehen zu können“ (Dyllick, 2018, S. 300). Eine dabei dominierende Methode stellt das Vor- und Nachsingen dar (Eberhard, 2016). Spychiger (2015) schreibt dieser Form des Lernens durch Imitation gerade im Primarstufenalter eine besondere Rolle zu.

Zudem kommen beim Singen weitere instruierende Handlungen der Lehrkraft ins Spiel, die insbesondere durch Mimik und Gestik vollzogen werden: Das Geben eines Einsatzes durch Dirigierbewegungen mit Händen und Armen oder ein deutlich sichtbares Einatmen der Lehrkraft als Atemimpuls für die Schüler*innen geben nicht nur Zeitpunkt, sondern auch Charakter des gemeinsamen Singens vor (z. B. laut oder leise, schnell oder langsam). Wenn solche Handlungen einer Lehrkraft – sei es im Kontext des Lernens einer neuen Melodie oder bei der Korrektur von Fehlern einer bekannten Melodie – primär auf den handelnden Nachvollzug der Schüler*innen ausgelegt sind, wird dies als Instruktion aufgefasst. Beim Vorsingen sollte sich die Lehrkraft dabei an den Adressat*innen, genauer am Klang der Kinderstimme (siehe z. B. Mohr, 2014), orientieren. Herausforderungen könnten sich in Singesituationen nach Eberhard (2016) gerade für Schüler*innen der Primarstufe ergeben, wenn eine männliche Lehrkraft mit nach unten oktavierter Singstimme vorsingt. Für das Unterrichtshandeln der Lehrkraft werden „berufsfeldspezifische musikpraktisch-künstlerische Fähigkeiten und Fertigkeiten“ (Puffer, 2021, S. 43) notwendig, die mit der Professionswissensfacette fachdidaktisches Wissen zusammenhängt – eine Verbindung zum Erklären und Instruieren ist auch innerhalb der Modellierung professioneller Kompetenzen von Musiklehrkräften sichtbar (siehe auch Puffer & Hofmann, 2022).

Instrumentales Musizieren im Unterricht kann entweder im normalen Unterricht oder in instrumentalen Profilklassen erfolgen, die ebenso wie das Konzept Chorklasse beispielsweise in Form von Streicher-, Bläser- oder Bandklassen realisiert werden (z. B. Heß, 2018a). In solchen Profilklassen, die meist im Teamteaching von einer Musik- sowie einer Instrumentallehrkraft durchgeführt werden, richten sich die Methoden vor allem nach den instrumentenspezifischen Charakteristika der Besetzung: Während im Bläserklassenunterricht zu Beginn beispielsweise häufig Atem- und Mundstückübungen im Zentrum stehen, startet der Streicherklassenunterricht häufig mit dem Zupfen der leeren Saiten (Heß, 2018a). Für das instrumentale Musizieren im *normalen* Musikunterricht mit Alltagsgegenständen oder Körperklängen liegen nicht nur „vielfältige Vorlagen von eigens komponierten Musikstücken“ (Eberhard, 2016, S. 62) vor, sondern es stehen je nach Ausstattung Boomwhackers, Stabspiele, Orff-Instrumentarium oder Keyboards in Klassenstärke zur Verfügung, die in den letzten Jahren durch digitale Medien und Technologien ergänzt werden (Godau, 2018). Auch wenn vermehrt nach Wegen eines eigenverantwortlichen und selbstbestimmten Lernens in Bezug auf Singen und Musizieren gesucht wird (Dyllick, 2018; Heß, 2018a), handelt die Lehrkraft ebenso wie beim Singen traditionell in eher frontalen Unterrichtssituationen (Fuchs, 2015), bei denen das Geschehen „in der Regel stark lehrer- bzw. leiterinnenzentriert“ (Puffer, 2021, S. 46) ist. Allerdings kann aufgrund des Einsatzes von Instrumentarium angenommen werden, dass eine Lehrkraft in einer konkreten Musiziersituation beispielsweise auch kurz erklärt, wie der

Klang eines bestimmten Instruments entsteht und worauf zu achten ist, wenn eine hohe Klangqualität erzielt werden soll (Klein, 2009).

Gerade bei der Vermittlung klanglicher Besonderheiten ausgewählter Instrumente werden auch Mischformen deutlich, bei denen eine Lehrkraft in einer Sing- oder Musizierphase nicht nur instruierend, sondern auch erklärend tätig sein kann. Dass jedoch primär instruierendes Handeln geboten ist und eine Lehrkraft möglichst wenig erklären soll, empfiehlt Eberhard (2010): In seiner Studie zu Unterrichtsstörungen im Musikunterricht der Realschule äußern Schüler*innen auf die Frage nach möglichen Gründen beispielsweise, dass Störungen dann auftreten, wenn Lehrkräfte in Musizierphasen „zu viel unterbrechen“ oder „zu viel erklären“ (S. 189). Da beim Singen und Musizieren das „Ideal eines länger anhaltenden bzw. über längere Zeit geführten musikalischen ‚Flusses‘ ohne Unterbrechung“ (Puffer, 2019, S. 233) verfolgt wird, sind – wenn überhaupt – nur kurze, den Musizierprozess ergänzende, Erklärungen sinnvoll, beispielsweise hinsichtlich Funktion und Physiologie der Singstimme oder bezüglich der Klangerzeugung beziehungsweise Spieltechnik eines Instruments (Kranefeld, Busch et al., 2015).

(2) Tanzen und sich zur Musik bewegen

Die folgende Darstellung folgt der Unterscheidung nach Vogel (2018), die motorische Bewegung in Bezug auf musikalisches Handeln in (a) „Bewegungen zur Klangerzeugung beim Instrumentalspiel“, (b) „Bewegung als Mittel zum Verständnis von Musik“ und (c) „Bewegungen, die als Tanz bezeichnet werden können“ (S. 334) unterteilt. Der Fokus richtet sich auf (b) und (c), während (a) für die folgende Darstellung vernachlässigt wird; diese Form von Bewegungen spielt vor allem im Instrumentalunterricht oder beim Klassenmusizieren in instrumentalen Profilklassen eine Rolle und wurde oben bereits entfaltet. Vogel (2018) definiert (c) Tanz als Form künstlerischen Ausdrucks, wenn dieser „selbstreferentiell“ (S. 334) und durch vorgegebene Bewegungs- und Schrittfolgen charakterisiert ist; außerdem dienen Tanzbewegungen „nicht dem Erreichen von Produkten außerhalb der Bewegung“, sondern liegen im „Vollzug der Bewegung“ (S. 334). Mit Blick auf das Handeln von Lehrkräften dominiert im Kontext von Tanzimprovisationen vor allem das Instruieren (z. B. Stibi, 2018 und 2019), während das Erklären eine eher untergeordnete Rolle zu spielen scheint. Nach Stibi (2019) äußert sich das beispielsweise dadurch, dass „die Lehrpersonen in ihren Instruktionen kaum Äußerungen über Charakter, Form und Struktur oder Aussage der gewählten Musikbeispiele machen“ (S. 259). Vielmehr hat die Lehrkraft bei der Tanzimprovisation das selbstgesteuerte Handeln der Schüler*innen im Blick beziehungsweise möchte sie dazu befähigen (siehe Hohenstein, 2009).

Wird jedoch von (b) Bewegung als Form künstlerischen Ausdrucks gesprochen, die insbesondere zum Verständnis von Musik beitragen soll, liegt nach Vogel (2018) „die Aufmerksamkeit sowohl auf dem Produkt als auch auf dem Prozess der Bewegungen selbst“ (S. 334). Im Musikunterricht können solche Bewegungsformen beispielsweise eine Rolle spielen, wenn Schüler*innen zu einem Hörbeispiel mitdirigieren und ihnen dadurch eine bewusste Wahrnehmung der metrischen Struktur eines Werks ermöglicht wird (Eberhard, 2016). Außerdem können Parameter, wie Tonhöhen oder

Melodieverläufe durch Bewegung „fein- oder grobmotorisch dargestellt werden“ (Vogel, 2018, S. 335). Musikalische Formen wie Rondo oder Suite sind ebenfalls durch verschiedene Bewegungen beziehungsweise Tanzformen erfahrbar (z. B. Buschendorff, 2010).

Durch den Zweck, mittels Bewegungen Musik für Schüler*innen verständlich(er) zu machen, rücken kognitive Ziele in den Blick (Steffen-Wittek, 2021). Bei diesen ist eine größere Anzahl an Situationen zu vermuten, in denen eine Lehrkraft etwas erklären muss – etwa bei der Weiterverarbeitung von einer mit der Hand grob nachgezeichneten Melodie hin zur konkreten Notation. Instruktionale Handlungen der Lehrkraft sind denkbar, wenn sie die Schüler*innen zum Mitdirigieren anleitet. Auch hier zeigt sich, dass beide Handlungen einerseits eng aufeinander folgen und dabei durchaus fließend ineinander übergehen, andererseits auf theoretischer Ebene getrennt voneinander klassifiziert werden können: Je nach verfolgtem Zweck dient das Instruieren beim Anleiten von Dirigierbewegungen primär der Befähigung zum Handlungsvollzug der Schüler*innen, das Erklären zum Aufbau von Wissen, wenn grobe Melodiestrukturen konkret notiert werden.

Zusammenfassend kann Erklären insbesondere dann eine Rolle spielen, wenn Bewegungen zur Musik ausgeführt werden, um musikalische Parameter oder Strukturen zu veranschaulichen. Im Gegensatz dazu spielt Erklären bei Unterrichtssituationen, in denen nach dem Begriffsverständnis Vogels (2018) von Tanz gesprochen werden kann, eine untergeordnete Rolle. Im Vordergrund steht instruierendes Unterrichtshandeln, bei dem lediglich kurze Erklärungen beispielsweise dann denkbar sind, wenn eine Lehrkraft Herkunft, Entwicklungsprozess oder Varianten eines Tanzes thematisiert. Es besteht die Annahme, dass Erklären während einer Tanzphase – analog zu den vorhin beschriebenen Sing- und Musizierphasen (siehe Eberhard, 2010) – keine relevante Rolle spielt und eine Lehrkraft vor allem instruierend tätig ist.

(3) Improvisieren und Komponieren von Musik

Innerhalb der musikpädagogischen Literatur wird kontrovers diskutiert, inwieweit Improvisieren und Komponieren miteinander zusammenhängen oder ob beide Handlungen eher getrennt voneinander betrachtet werden sollten. Während Krämer (2018a) eine explizite Trennung empfiehlt, da beispielsweise „das unmittelbare Agieren in der Zeit“ (S. 321) andere Kompetenzen als das Komponieren verlangt, betont Schlothfeldt (2018) im Gegensatz dazu aus didaktischer Perspektive, dass beide Handlungen nicht strikt voneinander getrennt werden sollten: Gerade bei Kindern seien „Improvisieren und Klangerkundung oft unabdingbare Bestandteile des Komponierens“ (S. 329), weshalb der Autor für das didaktische Vorgehen im Musikunterricht das Übergehen von Improvisieren zum Komponieren empfiehlt. Um das Komponieren im unterrichtlichen Kontext von der professionellen Tätigkeit zu unterscheiden, schlägt Schlothfeldt (2018) vor, sich etablierter Begriffe wie *Musik erfinden* (z. B. Jeismann & Kranefeld, 2021; Kranefeld et al., 2019; Nimczik, 1997) oder *Produktion* (z. B. Wallbaum, 2009) zu bedienen. Für die nachfolgende Darstellung sind demnach schöpferische Prozesse relevant, in denen Musik sowohl in Improvisations- als auch Kompositionsprozessen erfunden wird.

In Bezug auf das Unterrichtshandeln einer Lehrkraft spricht Krämer (2018a) von der Aufgabe, „eine Gruppe von Improvisierenden anzuleiten“ (S. 325). Er empfiehlt

diesbezüglich, vor der Improvisation Herausforderungen zu formulieren und im Nachgang eine Reflexion der Schüler*innen zu moderieren (Krämer, 2018a). Dabei schreibt der Autor einer Lehrkraft keine leitende Funktion im Sinne des Dirigierens zu, sondern sieht sie eher als „primus inter pares“, wodurch der „sozial-kommunikative Aspekt und die musikalische Interaktion in der Gruppe“ (Krämer, 2018a, S. 325) stärker in den Vordergrund gerückt wird. Nach Krämer (2018a) soll eine Musiklehrkraft demnach während eines solchen Schaffensprozesses möglichst weder erklären noch instruieren. Allerdings können solchen Improvisationsphasen instruierende Aufgabenstellungen vorausgehen – etwa beim Improvisieren im pentatonischen Raum (siehe z. B. Jank & Schmidt-Oberländer, 2010, S. 202–212). Nach Lang (2010) ist es zudem möglich, dass eine Lehrkraft Improvisationsprozesse auch durch kurze Erklärungen fruchtbar vorbereiten kann: Der Autor untersuchte bei Schüler*innen einer 7. Klasse das Lernpotenzial eines kurzen musiktheoretischen Inputs hinsichtlich der Entwicklung von Improvisationsmustern auf einem Tasteninstrument und kann in vielen Fällen eine „Weiterentwicklung der persönlichen Spielweise“ (Lang, 2010, S. 271) beobachten.

Auch können kurze Erklärungen und Instruktionen Kompositionsprozesse einleiten. Dabei dienen außermusikalische Bezugspunkte durch „die Hinzunahme von bildnerischen, literarischen oder szenischen Vorlagen“ als Grundlage für das Erfinden von Musik (Kranefeld, 2008, S. 77). Konkrete Aufgabenstellungen können dabei einen instruierenden Charakter aufweisen, indem Handlungsschritte vorgegeben werden. Wenn durch sie notwendige Bezüge zwischen außermusikalischem Bezugspunkt und der zu erstellenden Komposition hergestellt werden, übernehmen sie auch erklärende Funktion. Umgekehrt kann auch versucht werden, musikalische Strukturen tiefer zu durchdringen, indem Schüler*innen Melodieverläufe bildnerisch darstellen und auf einem Blatt nachzeichnen oder Musik in Form von Gedichten oder Geschichten versprachlichen (Krämer, 2018b). Hier können die erschaffenen Produkte im Nachhinein zum Verständnis musikalischer Strukturen beitragen und damit auch erklärende Funktion einnehmen. Im Gegensatz zu den ersten beiden Handlungsfeldern, in denen Erklären und Instruieren vermehrt während des musikalischen Prozesses zum Einsatz kommen kann, flankieren beide Handlungen schöpferische Prozesse im Erfinden von Musik und treten eher im Voraus oder im Nachgang auf.

(4) Wahrnehmen von Musik

Lauer (2018) konstatiert, dass sich der Terminus Musik Wahrnehmen gegenüber der Bezeichnung Musik hören im Laufe der Zeit immer mehr durchgesetzt zu haben scheint. Da Weidner (2018b) zwischen den zwei Begriffen jedoch eine Art Koexistenz feststellt und es nahezu unmöglich scheint, den einen ohne den anderen zu verhandeln, bezieht die folgende Darstellung beide mit ein.

In Bezug auf Musikhören unterscheidet Weidner (2018b) zunächst „zwischen eher technischen und auf einzelne Strukturelemente bezogenen Ansätzen zur *Gehörbildung*“ und einer „umfassender angelegten Hörerziehung“ (S. 315). Während Konzepte von einer kognitiv ausgerichteten Gehörbildung in der heutigen Zeit vor allem in verschiedenen Studiengängen des Fachs Musik im Bereich Musiktheorie eine Rolle spielen, wurden sie

im schulischen Kontext zugunsten einer „Ausbildung ganzheitlich gedachter Zugänge zum Hören“ (Weidner, 2018a, S. 317) in den letzten Jahrzehnten⁵⁶ immer mehr verdrängt. Für ein Beispiel zur Veranschaulichung von Erklären und Instruieren werden Inhalte des Lehrplans für musikalisches Gymnasium in Bayern herangezogen: Schüler*innen sollen bereits in der fünften Jahrgangsstufe in der Lage sein, kurze rhythmische Folgen hörend zu erfassen und sie im Anschluss zu notieren (Mu5 Lernbereich 5: Musik und ihre Grundlagen, siehe Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2017a). Aufgabe der Lehrkraft ist dabei nicht nur, Strategien zu vermitteln, sondern auch zu erklären, wie diese erfolgreich anzuwenden sind. Für ein Rhythmusdiktat im 4/4-Takt kann eine solche Strategie lauten: Füge entsprechend der Anzahl an Schlägen pro Takt (Anzahl von 4 Schlägen im Wert einer Viertelnote) kleine Kästchen unterhalb der Notenlinie hinzu, um den Rhythmus in kleinere Einheiten strukturieren zu können. Neben dem Instruieren der Handlungsschritte sollte die Lehrkraft beispielsweise erklären, in welchem Verhältnis die Anzahl der Kästchen zur jeweiligen Taktart steht.

Der Begriff Wahrnehmen geht über Ansätze der Gehörbildung oder Hörerziehung hinaus und beinhaltet eine multisensorielle Herangehensweise, die sowohl „emotional-affektiv als auch intellektuell bzw. kognitiv orientierte Rezeptionsweisen von Musik“ (Weidner, 2018a, S. 315) einschließt. Die Entwicklung von Wahrnehmungsfähigkeit fokussiert dabei nicht nur die „Ausbildung des Gehörs hinsichtlich Differenzierungsfähigkeit und Gedächtnisleistung“, sondern bezeichnet neben einer „physiologischen Wahrnehmung äußerer Reize („perception“),“ auch deren „erkennende Verarbeitung („cognition“),“ sowie das Übersetzen in Klang (z. B. beim Nachsingen), Grafik, Verbalsprache oder Schrift“ (Weidner, 2018a, S. 315).

Diese Art der Konzeptualisierung zeigt sich in den vier Dimensionen des Kompetenzmodells *Musik wahrnehmen und kontextualisieren* (z. B. Jordan et al., 2012; Knigge, 2011; Niessen et al., 2008): Gedächtnisfähigkeit auf der einen Seite und Kompetenzen im Notieren, Verbalisieren sowie Einordnen von Musik in verschiedene historische wie kulturelle Kontexte auf der anderen Seite. Gerade bei der Unterstützung solcher kognitiven Vorgänge, wie etwa das angesprochene Weiterverarbeiten äußerer Reize und die Überführung in Notation oder Fachsprache, können Lehrkräfte Zusammenhänge aufzeigen oder Notationsverfahren sowie die Bedeutung einzelner Fachbegriffe erklären. Deutlich wird hier auch die enge Verknüpfung bis hin zur Verschränkung mit den Bereichen Produktion, Reproduktion, Transformation beziehungsweise Reflexion, die dem Wahrnehmen von Musik eine Art Meta-Position in Bezug auf die anderen Handlungsfelder zuweist (Laufer, 2018).

Es ist anzunehmen, dass im Kontext der Dimension *Musik wahrnehmen* insbesondere aufgrund der engen Verbindung zu anwendungsbezogenen Kontexten wie Notieren oder Verbalisieren von Musik Zusammenhänge aufgezeigt und neue Wissensaspekte in vorhandene integriert werden müssen. Hinsichtlich dieses Zwecks ist zu erwarten, dass eine Musiklehrkraft vermehrt etwas erklären muss, während instruierende Handlungen eher im Kontext der Strategievermittlung eine Rolle spielen können.

56 Für einen historischen Überblick der zentralen Konzeptionen zum Musikhören siehe Weidner (2018b).

(5) Nachdenken und Sprechen über Musik

Nach Weidner (2018c) umfasst das didaktische Handlungsfeld Reflexion das Nachdenken und Sprechen über sowie auch das Analysieren von Musik, wobei die Autorin aufgrund vielfältiger Phänomene und Perspektiven des Reflexionsbegriffs „auf eine vereinheitlichende Definition [...] verzichtet“ (S. 348). In diesem Kontext wird häufig die Vermittlung von Inhalten aus den Bereichen Musiktheorie, Musikgeschichte und Instrumentenkunde genannt (z. B. Biegholdt, 2013; Eberhard, 2010). Unterschieden wird bei musiktheoretischen Inhalten zwischen dem stark analyseorientierten und anwendungsorientierten Bereich der Musiktheorie und konkreten Wissensinhalten über Notenwerte, Akkorde, Taktarten oder Tonleitern, die unter dem Begriff Allgemeine Musiklehre zusammengefasst werden (Weidner, 2015).

Während Reflexion aus Sicht von Biegholdt (2013) alle anderen Handlungsfelder durchdringt und „jeweils immer im Zusammenhang mit einer oder mehreren Umgangsweisen zu sehen ist“ (S. 51), zeigen verschiedene Ansätze, dass sich gerade das Zusammenspiel von Reflexion mit Hör- beziehungsweise Wahrnehmungsprozessen als erfolgversprechend erweist. Beide Handlungsfelder waren nicht nur zentrale Säulen einflussreicher musikdidaktischer Konzeptionen (siehe z. B. Alt, 1968; Ehrenforth, 1971; Richter, 1976; kritisch: z. B. Vogt, 2016), sie sind auch Gegenstand aktueller Modellierungen der Kompetenz *Musik wahrnehmen und kontextualisieren* (Jordan et al., 2012). Weidner (2018c) sieht dabei das Sprechen und Nachdenken über Musik als Verarbeitungsprozess, der nach der Wahrnehmung äußerer Reize in Gang gesetzt wird und sich in der Übersetzung in Verbalsprache und Schrift beziehungsweise Notation äußern kann. Damit steht Reflexion in enger Verbindung mit dem Wahrnehmungsbegriff, der im vorherigen Absatz erläutert wurde.

Während das Handlungsfeld Reflexion aufgrund des bereits dargelegten Paradigmenwechsels hin zu einer Handlungsorientierung im Musikunterricht seit den 1980er Jahren verstärkt in den Hintergrund gerückt ist (Jank, 2013; Krause-Benz, 2016), wird seine Bedeutung im aktuellen Diskurs unterschiedlich verhandelt: Auf der einen Seite findet das Handlungsfeld *Reflexion* in einzelnen musikdidaktischen Publikationen kaum mehr als eine kurze Erwähnung (z. B. Biegholdt, 2013)⁵⁷, auf der anderen Seite bezeichnet es Eberhard (2016) als „zweifelloso eine wichtige Säule des Musikunterrichts“ (S. 64) – an anderer Stelle wird sogar von einer „Renaissance des Redens über Musik“ (Oberhaus, 2015, S. 54) im Unterricht gesprochen. Insgesamt zeigt sich hinsichtlich der Bedeutung des Handlungsfelds Reflexion ein ambivalentes und viel diskutiertes Bild.

Mit Blick auf Erklären als Unterrichtshandeln einer Lehrkraft sind in den genannten Kontexten verschiedene Einsatzsituationen denkbar. So kann beispielsweise erklärt werden, unter welchen technischen und physikalischen Voraussetzungen bestimmte Instrumente oder Gegenstände Klänge produzieren. Hier sind – analog zum Singen und

57 Beispielsweise räumt Biegholdt (2013) in seiner musikdidaktischen Publikation *Musik unterrichten* der Darstellung der übrigen vier Umgangsweisen 18 Seiten ein, während er dem Bereich Reflexion fünf Zeilen widmet und lediglich darauf verweist, dass dieser alle anderen Bereiche durchdringt und „immer in Zusammenhang mit einer oder mehrerer der anderen Umgangsweisen zu sehen“ (S. 51) ist.

Musizieren – jedoch auch Situationen denkbar, in denen Schüler*innen zum eigenständigen Spielen von Instrumenten befähigt werden sollen. In diesem Kontext würde eine Lehrkraft instruieren, in dem sie beispielsweise beim Spielen eines E-Basses zum Nachmachen einer deutlich sichtbaren Zupfbewegung auffordert. Im Bereich Musikgeschichte können kurze Erklärungen hinsichtlich der Entstehung von Antonin Dvoraks 9. Sinfonie dann sinnvoll sein, wenn darin politische, gesellschaftliche und ästhetische Einflüsse aufgezeigt werden, die den Kompositionsprozess beeinflusst haben (siehe Schmid, 2019, S. 60–67); das Instruieren spielt in diesem Bereich eine untergeordnete Rolle.

Im Kontext des Sprechens und Nachdenkens über sowie Analysieren von Musik wird das Wissen über Zusammenhänge musiktheoretischer Inhalte als notwendig erachtet (z. B. Biegholdt, 2013; Eberhard, 2010). Deren Vermittlung wird in Kapitel 2.4.3 ausführlich und mit Blick auf unterrichtliches Erklären entfaltet. Dabei werden auch die eng verlaufenden Trennungslinien zwischen Erklären-Wie und Instruieren aufgegriffen, die im vorherigen Abschnitt offengelegt wurden.

Zusammenfassung

Zusammenfassend wird für die Unterscheidung von Erklären und Instruieren zweierlei deutlich: Zum einen können beide Handlungen einer Lehrkraft in konkreten Unterrichtssituationen anhand des Zwecks theoretisch voneinander unterschieden werden. In der praktischen Umsetzung müssen diese Unterschiede jedoch nicht zwingend sichtbar sein, was am Vorspielen von Instrumenten beispielhaft deutlich wurde: Je nach Unterrichtssetting und -ziel kann es instruierende oder erklärende Funktion einnehmen. Zum anderen können für Erklären und Instruieren Schwerpunkte hinsichtlich der einzelnen Lernbereiche ausgemacht werden. Für die Handlungsfelder *Singen und Musizieren* sowie *Tanzen und sich zur Musik bewegen* wurden Situationen skizziert, in denen eine Lehrkraft während des musikalischen Prozesses vor allem instruiert. Im Kontext der letzten beiden Handlungsfelder *Musik wahrnehmen* und *über Musik nachdenken und sprechen* kann insbesondere dem Erklären eine große Bedeutung beigemessen werden. Schöpferische Prozesse des *Musikerfindens* sollten sowohl von Instruktionen als auch Erklärungen nur flankiert werden und entweder im Voraus oder im Nachgang erfolgen.

Im Folgenden fokussiert die Darstellung ausschließlich unterrichtliches Erklären im Fach Musik und beleuchtet dabei die Vermittlung musiktheoretischer Inhalte aus dem Bereich Allgemeiner Musiklehre. Sie dienen in der vorliegenden Arbeit dazu, gutes Erklären im Musikunterricht zu konzeptualisieren, da ihnen einerseits ein besonders hoher Bedarf am Erklären attestiert werden kann. Andererseits erlaubt die kognitive Auseinandersetzung mit Fachinhalten gleichermaßen auch einen fachübergreifenden Anschluss hinsichtlich potenzieller Qualitätsmerkmale guten Erklärens.

2.4.3 Erklären musiktheoretischer Inhalte

Der Erwerb musiktheoretischer Kenntnisse wird dem für den Musikunterricht wichtigen Handlungsfeld Reflexion zugeordnet (z. B. Eberhard, 2016; Kaiser & Nolte, 1989;

Venus, 1969), das – ähnlich wie auch das Wahrnehmen beziehungsweise Hören (Reflexion) – querschnittartige Aufgaben übernimmt und in alle anderen Felder hineinwirkt (Biegholdt, 2013). In den vorherigen Ausführungen wurden bereits Situationen skizziert, in denen eine Lehrkraft musiktheoretische Inhalte erklärt. Dabei ist jedoch zu berücksichtigen, dass im musikpädagogischen Diskurs keine einheitliche Auffassung hinsichtlich der Rolle musiktheoretischer Inhalte im Musikunterricht ausgemacht werden kann: Auf der einen Seite wird über die Sinnhaftigkeit der Vermittlung musiktheoretischer Inhalte kontrovers diskutiert; auf der anderen Seite nehmen diese Inhalte einen seit langer Zeit und bis heute relevanten Auftrag innerhalb der curricularen Vorgaben für einen allgemeinbildenden Musikunterricht ein (Weidner, 2015). Diese Ambivalenz verdeutlicht Lang (2008) in einer Metapher: Die Musiktheorie⁵⁸ sei im Musikunterricht „wie ein ungeliebter Gast“ und werde zwar aus Angst vor ausbleibenden soliden Kenntnissen sowie „aus schlechtem Gewissen gegenüber dem Lehrplan“ (S. 85) eingeladen, wirklich stören würde ihre Abwesenheit aber auch nicht. Dies könnte ein möglicher Grund dafür sein, dass das Erklären musiktheoretischer Inhalte im musikpädagogischen Diskurs kaum eine Rolle spielt.

Nachdem im Folgenden zunächst geklärt wird, was unter dem Terminus *musiktheoretische Inhalte* zu verstehen ist und wie sich dieser von *Musiktheorie* unterscheidet, sind anschließend vor allem zwei Fragen bedeutsam, die in dieser Reihenfolge beantwortet werden. Erstens: *Warum* sollen musiktheoretische Inhalte im Unterricht vermittelt werden? Und zweitens: *Wie* sollen die Inhalte vermittelt werden, beziehungsweise welche Rolle fällt diesbezüglich dem Erklären zu?

Terminologische Rahmung

Sowohl die Terminologie *musiktheoretische Inhalte* als auch der Begriff *Musiktheorie* können im Kontext von schulischem Musikunterricht nicht ohne Erläuterung verwendet werden. Es stehen beispielsweise die Fragen im Raum, in welchem Verhältnis Musiktheorie zu Musikpraxis steht oder welche Gegenstände der Begriff *musiktheoretische Inhalte* genau umfasst. Im Folgenden wird begründet, warum in der vorliegenden Arbeit grundsätzlich von musiktheoretischen Inhalten gesprochen wird.

Zunächst kann beobachtet werden, dass der Begriff *Musiktheorie* im musikpädagogischen Diskurs häufiger als *musiktheoretische Inhalte* verwendet wird und weitaus geläufiger ist. Allerdings lässt die Forschungsliteratur eine einheitliche Definition vermissen. Kühn (2002) legt bereits verschiedene Ebenen offen, indem er „Musiktheorie als autonome, für sich denkende Theorie“ von „Musiktheorie als pädagogische[s], auf andere gerichtete[s] Fach“ (S. 11) unterscheidet. Eine erste Beschränkung ist vor dem Hintergrund von schulischem Musikunterricht demnach auf pädagogische Kontexte vorzunehmen, wobei im weiteren Verlauf auch noch dezidiert zwischen Schule und Hochschule

58 Der Begriff *Musiktheorie* wird unterschiedlich verwendet (Weidner, 2015). Hinsichtlich seiner Bedeutung und dem Verhältnis zur Terminologie *musiktheoretische Unterrichtsinhalte* erfolgt eine Konkretisierung für die vorliegende Arbeit im weiteren Verlauf des Teilkapitels 2.4.3.

unterschieden wird. Wenn *Musiktheorie* jedoch in der vorliegenden Arbeit gebraucht wird, meint der Begriff immer eine Menge an Inhalten, wie sie im vorherigen Absatz (siehe Kap. 2.4.2, S. 109–111) aufgezählt wurden. Ruf (2014) arbeitet einen solchen „materialen Begriff von Musiktheorie“ (S. 136) auch in den Äußerungen von Musiklehrkräften hinsichtlich des Unterrichtens von *Musiktheorie* heraus. In diesem Kontext wird jedoch auch eine weitere unscharfe Begriffsbeziehung deutlich: Eine Lehrkraft äußert zur Frage, was „für sie ‚im Musikunterricht alles zum Komplex Theorie‘ gehöre“, dass beispielsweise „Anhören (...) ja eigentlich auch keine Praxis sei“ und unter „Theorie“ generell das zusammengefasst werden könne, „was die Schüler nicht selbst praktisch machen“ (Ruf, 2014, S. 135). In dieser Aussage wird – allerdings auch provoziert von der Fragestellung – zunächst ein Verständnis deutlich, in dem sich die Begriffe Musiktheorie und Musikpraxis gegenseitig ausschließen und unter Musikpraxis beispielsweise das zu fassen ist, was in singender, sich bewegender oder musizierender Weise ausgeführt wird. Im weiteren Verlauf schränkt die Lehrkraft in ihrer Aussage allerdings ein, dass sie unter *Musiktheorie* vor allem die Beschäftigung am Notenbild meine (Ruf, 2014).

Deutlich wird, dass der Begriff Musiktheorie – beispielsweise aufgrund eines unklaren Verhältnisses zum Begriff Musikpraxis – unterschiedlich aufgefasst wird und seine Bedeutung insgesamt unscharf ist. Um dies klarer zu fokussieren, wird in den folgenden Ausführungen aufgezeigt, warum der Terminus *musiktheoretische Inhalte* für die vorliegende Arbeit geeignet erscheint.

Weidner (2015) zählt zu musiktheoretischen Inhalten, die weitgehend einheitlich in älteren wie neueren Schulbüchern thematisiert werden (z. B. Kimmelmeyer & Nykrin, 2004; Prinz & Scheyett, 1993; Schmid et al., 2019), „zum einen rhythmisch-metrische Grundelemente wie Notenwerte und Taktarten, zum anderen tonhöhenbezogene Strukturierungen wie ausgewählte, im Violinschlüssel notierte Töne, verschiedene Intervalle und Tonleitern sowie akkordische Strukturen und einfache harmonisch-kadenzielle Verbindungen“ (Weidner, 2015, S. 133). Darüber hinaus nimmt die Autorin in ihrer Forschungsarbeit, in der sie die Beziehung der beiden Fächer Musikpädagogik und Musiktheorie aus einer systemtheoretischen Perspektive heraus untersucht, weitere terminologische Konkretisierungen vor: Solche musiktheoretischen Inhalte, die in musikalischen Schulcurricula aufgeführt sind, können nur unzureichend unter dem Begriff *Musiktheorie* gefasst werden, da dieser eher im Kontext der Hochschullehre zu verorten ist und dort für das akademische Fach steht, das verschiedene Bereiche wie Gehörbildung oder Tonsatz zusammenfasst (Weidner, 2015; siehe auch Richter, 2002; Sprau, 2017).

Da diese Inhalte weit über das hinausgehen, was als Kern der Vermittlung musiktheoretischer Inhalte – gerade im Bereich der Primarstufe sowie in der Sekundarstufe I – beschrieben ist, werden daher im Kontext schulischen Musikunterrichts einerseits begriffliche Konkretisierungen wie „schulische Musiktheorie“ (Weidner, 2012, S. 305) vorgenommen oder es begegnen andererseits Begriffe wie „Allgemeine Musiklehre“ (Fuß, 1997, S. 153), „elementare Musiklehre“ (Ruf, 2014, S. 137) oder „Elementarlehre“ (Lang, 2008, S. 94). Die vorliegende Arbeit meint musiktheoretische Inhalte, die sowohl von Weidner (2015) eingangs als solche klassifiziert wurden als auch an anderer Stelle unter dem Begriff *Allgemeine Musiklehre* zusammenfasst werden (Fuß, 1997). Nach

Weber (2009) entsprechen diese Inhalte einer praktischen Sprachlehre beziehungsweise Grammatik, während der Begriff *Musiktheorie* die Ebene der wissenschaftlichen Sprachlehre darstellt und größtenteils das Curriculum von Hochschulstudiengängen beziehungsweise in Ansätzen auch schulischem Musikunterricht der Sekundarstufe II beschreibt.

Musiktheoretische Inhalte im Unterricht

Die nun zu beantwortende Frage, warum musiktheoretische Inhalte im Musikunterricht vermittelt werden sollen, ist verbunden mit dem Diskurs um musikalische Bildung, beziehungsweise der Frage danach, ob musiktheoretische Wissensinhalte als Teil musikbezogener Kompetenz einen Beitrag zu musikalischer Bildung leisten. Nach Vogt (2012) sei unstrittig, „dass zur musikalischen Bildung erlernbare musikalische ‚Kompetenzen‘ notwendig sind; ebenso unstrittig ist aber, dass musikalische Bildung nicht in (...) musikbezogenen Fähigkeiten und Fertigkeiten aufgeht“ (S. 19).⁵⁹ Die Antwort auf die Frage nach Gründen der Vermittlung musiktheoretischer Inhalte erfolgt anhand unterschiedlicher Perspektiven.

Eine erste Antwort ist innerhalb der Bildungswissenschaft zu finden (siehe z. B. Tenorth, 2004). Dort wird die Frage nach Bildungsstandards, beziehungsweise danach, welchen Beitrag die jeweiligen Schulfächer leisten können, aus bildungstheoretischer wie empirischer Perspektive diskutiert. Tenorth (2004) schlägt dabei vor, „die Diskussion um Bildungsstandards in den Kontext der kulturwissenschaftlichen Debatte um das kulturelle Gedächtnis und Prozesse der Kanonisierung“ (S. 650) zu stellen, wobei er in diesem Zusammenhang den Begriff „kultureller Literalität“ (S. 654) prägt. Deutlich wird diese Kontextualisierung beim Blick auf die PISA-Studie. Das ihr zugrundeliegende *literacy*-Konzept versteht Welterschließung als Fähigkeit, „die Welt auf unterschiedliche Weise lesen zu können“ und ist „an den spezifischen Rationalitätsmustern der schulischen Fächer und ihrer jeweiligen wissenschaftlichen ‚Domänen‘“ (Dressler, 2007, S. 250) ausgerichtet.

Demnach gibt es unterschiedliche Zugänge zur Welt, die Baumert et al. (2001) als „Modi der Welterfahrung“ bezeichnen und als „Orientierungswissen vermittelnde Begegnung mit kognitiver, moralisch-evaluativer, ästhetisch-expressiver und religiös-konstitutiver Rationalität“ (S. 21) auffassen. Diese vier Zugänge sind als jeweils eigene Lesart der Welt zu verstehen und nach Dressler (2007) weder hierarchisch zu denken noch „wechselseitig substituierbar“ (S. 250). Musik beziehungsweise das Schulfach Musik ermöglicht nach Baumert (2002) eine „ästhetisch-expressive Begegnung und Gestaltung“ (S. 113) der Welt.⁶⁰ Gleichzeitig verlangt ein damit verbundenes „ästhetisches und expressives Verstehen“ (Tenorth, 2004, S. 654) der Welt nach Kompetenzen, um die

59 Welche Kompetenzen das sein können, bleibt unklar. In Bezug auf den Mangel an Bildungsstandards für das Fach Musik (siehe z. B. Hasselhorn & Knigge, 2018; Heß, 2018b; Jank et al., 2020) gibt es beispielsweise Vorschläge in Form musikpraktischer Kompetenzen (z. B. Hasselhorn & Lehmann, 2014) sowie im Bereich *Musik wahrnehmen und kontextualisieren* (z. B. Jordan et al., 2012).

60 Während diesem Weltzugang neben Musik beispielsweise noch die sprachlich-literarischen Fächer sowie Kunst und Sport zugeordnet werden, gestalten sich nach Baumert (2002) die drei anderen Zugänge als „Kognitiv-instrumentelle Modellierung der Welt“, „Normativ-evaluative Auseinandersetzung mit Wirtschaft und Gesellschaft“ und „Probleme konstitutiver Rationalität“ (S. 113).

jeweilige *Sprache* lesen und verstehen zu können. Als eine Form ästhetischen Verstehens ist also auch Musik als Sprache beziehungsweise Form der Kommunikation aufzufassen.

Eine dazu ähnlich verlaufende, musikpädagogische Perspektive nimmt Richter (2002) ein, der es als eine Aufgabe von Musikunterricht ansieht, „Musik, verstanden als spezieller Bereich menschlicher Wirklichkeit [...] zu verklären, zu verdeutlichen und ihr Verstehen zu eröffnen – in historischer Hinsicht, im Hinblick auf andere Kulturen und im Hinblick auf allgemeine Grunderfahrungen“ (S. 48). Dabei sei es nach Meinung des Autors unerlässlich, sich unter anderem mit Fragen zu beschäftigen, „auf welchen Materialien [...] Musik und ihre Formung beruht“ oder „welche Muster und Regeln in ihr zu entdecken sind“ (Richter, 2002, S. 48). Ein derart geprägtes Wissen als Bestandteil einer *musikbezogenen Literalität* bedeutet vor dem Hintergrund, dass Reflexion und damit auch das Wissen um musiktheoretische Inhalte in alle anderen Umgangsweisen hineinragen: Wer über Musik sprechen, Musik reproduzieren oder Musik komponieren und für andere nachvollziehbar notieren⁶¹ möchte, benötigt dafür nicht nur Wissen über das einheitlich verwendete und allgemein verständliche Zeichensystem sowie Vokabular. Ein anwendungspraktischer Umgang – wie es Dressler (2007) nennt – erfordert auch Wissen über tonale, rhythmische und metrische Zusammenhänge beziehungsweise *Materialien*.

Außerdem kann die Musikpädagogik bezüglich der Vermittlung musiktheoretischer Inhalte eine seit vielen Jahrhunderten gewachsene und in verschiedenen Kulturen verbreitete Tradition vorweisen, deren Historie an dieser Stelle knapp skizziert wird. Meilensteine waren diesbezüglich etwa die im 11. Jahrhundert von Guido von Arezzo entwickelte Notenschrift sowie die in diesem Zusammenhang eingeführte Technik der Solmisation, die heute in verschiedenen Kulturkreisen als Methode für die Vermittlung musiktheoretischer Inhalte eingesetzt wird (Ehrenforth, 2010). Ehrenforth (2010) berichtet zudem, dass sich in den nachfolgenden Jahrhunderten ein zweigleisiges musikalisches „Bildungsangebot in den avancierten Kloster-, Dom- und Stiftsschulen“ (S. 180) herausgebildet hat, welches sowohl die Hinführung zur *musica practica* (z. B. gottesdienstliches Singen) als auch zur *musica theoretica* (z. B. Lesen einschlägiger musiktheoretischer, theologischer wie philosophisch-politischer Werke) zum Ziel hatte.

Diese grundsätzliche Unterscheidung in eine musizierende und eine analytische Auseinandersetzung mit Musik hat bis heute hinein ebenso Gültigkeit, wie die Frage nach Zusammenhängen und Überschneidungen beider Bereiche (Lang, 2008). Die Zeit um das 18. Jahrhundert war begleitet von einem Mangel an flächendeckender institutioneller musikalischer Ausbildung, weshalb dieser durch eine Vielzahl an Publikationen musikalischer Lehrwerke auszugleichen versucht wurde; diese richteten sich an Lehrende wie Lernende und setzten musiktheoretische Inhalte für die Arbeit mit dem Buch entweder voraus oder behandelten sie wie Leopold Mozarts Lehrwerk *Versuch einer gründlichen Violinschule* ausführlich (Ehrenforth, 2010).

Flächendeckend institutionalisiert wurde die Vermittlung musiktheoretischer Inhalte in Deutschland vor allem durch die in der Kestenbergs-Reform geschaffenen Neuordnungen zum Musikunterricht in der Schule und der damit verbundenen Ausbildungsstruktur

61 *Literalität* kann also hier verstanden werden als Lese- und Schreibkompetenz hinsichtlich der Sprache Musik.

für Lehrkräfte beziehungsweise der Organisation von Instrumental- sowie Gesangsunterricht an Musikschulen (Gruhn, 2003a). Gerade die Neuausrichtung schulischen Musikunterrichts bildete die Grundlage für musikdidaktische Konzeptionen, die bei der Darstellung von Erklären und Instruieren im Lichte musikdidaktischer Handlungsfelder teilweise vorgestellt wurden (siehe Kap. 2.4.2). In der historischen Rückschau zeigt sich die bedeutende Rolle der Vermittlung musiktheoretischer Inhalte, die sich bis heute auch in den Curricula der allgemeinbildenden Schulen niederschlägt.

Dabei wird deutlich, dass die Struktur der Lehrpläne im Fach Musik größtenteils durch die Orientierung an Umgangsweisen mit Musik zu erklären ist (Kaiser & Nolte, 1989). Musiktheoretische Inhalte finden sich sowohl in den Curricula der Primar- als auch der Sekundarstufe wieder, wobei die bereits dargelegten Inhalte wie Notenwerte, Taktarten, Intervalle oder Tonleitern „nahezu überall präsentiert“ (Weidner, 2015, S. 133) werden.⁶² Mit Blick auf die in den letzten 20 Jahren gültigen bayerischen Lehrpläne des Gymnasiums im Fach Musik können Änderungen hinsichtlich der Vermittlung musiktheoretischer Inhalte festgestellt werden. Schäfer-Lembeck (2003) stellt über den seinerzeit gültigen bayerischen Lehrplan für Gymnasien fest, dass es wie ein Gebot zu verstehen ist, selbstzweckhaftes Unterrichten von Musiktheorie abzulehnen (siehe auch Beiderwieden, 2008; Jank, 2013) und musiktheoretische Inhalte nur in Verbindung mit musikpraktischen Anteilen zu thematisieren. Musikpraxis stehe dabei grundsätzlich an erster Stelle des Unterrichts (Schäfer-Lembeck, 2003).

Im danach folgenden und aktuell gültigen LehrplanPLUS wird *Musiktheorie*⁶³ als einer von vier Gegenstandsbereichen benannt, allerdings bleibt nach Sprau (2017) unbestimmt, welche Inhalte sich dahinter verbergen – mit Blick auf die Kompetenzerwartungen werden jedoch jene, dort auch so bezeichnete, Fachinhalte wie Notenwerte oder Taktarten deutlich, die weiter oben unter dem Begriff Allgemeine Musiklehre subsummiert wurden (siehe Weidner, 2015). Die Kombination musiktheoretischer und musikpraktischer Elemente, wie sie der vorherige Lehrplan für den Musikunterricht noch explizit forderte, wird auch im LehrplanPLUS angesprochen – dort jedoch eher implizit. So ist etwa als Kompetenzerwartung formuliert: „Die Schüler*innen beschreiben einfache Melodie- und Rhythmusverläufe und vollziehen sie praktisch nach“ (Mu5, Lernbereich 4: Musik und ihre Grundlagen, Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2017a). Nach Sprau (2017) bleibt diesbezüglich jedoch unklar, ob den Lernenden nicht weiterhin vor allem Fachinhalte an Stelle der anvisierten Kompetenzen vermittelt werden, weshalb weiterhin von der Gefahr auszugehen ist, dass „schulische Leistungserhebungen auch in Zukunft die Bereitstellung von Inhalten anstelle des kompetenten Umgangs mit ihnen abprüfen werden“ und dabei das „Potenzial

62 Als exemplarische Beispiele dienen die Lehrpläne aus Bayern (Lehrplanbereich Musik und ihre Grundlagen in Gymnasium: Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2017a; Realschule: Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2017b) und Nordrhein-Westfalen (Ordnungssysteme musikalischer Strukturen in Gymnasium, Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2019).

63 Beispielsweise für die 5. Jahrgangsstufe des Gymnasiums in Bayern: Mu5 – Lernbereich 4: Musik und ihre Grundlagen (siehe Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2017a).

musiktheoretischer Herangehensweisen an den musikalischen Gegenstand auf diese Weise größtenteils ungenutzt bliebe“ (S. 130).

Diese Befürchtung spiegelt sich auch in Vorstellungen von Musiklehrkräften bayerischer Gymnasien über das Lehren und Lernen musiktheoretischer Inhalte wider (siehe Ruf, 2014). Aus Sicht der Lehrkräfte sind musiktheoretische Kenntnisse für die Erschließung von Musik notwendig, allerdings wird das Unterrichten musiktheoretischer Inhalte auch damit begründet, curricularen Vorgaben und der Vergabe von Zensuren nachzukommen (Ruf, 2014; siehe auch Beiderwieden, 2008; Lang, 2008 und 2010). Hinsichtlich der Ziele, die Musikunterricht verfolgen sollte, werden „zum einen Spaß, Freude, Interesse und Bereicherung an und durch Musik genannt, zum anderen kognitive Ziele“ (Ruf, 2014, S. 138) – beides zu vereinbaren wird jedoch nicht nur als herausfordernd und schwer umsetzbar, sondern an manchen Stellen sogar als unmöglich empfunden, da das Verfolgen kognitiver Ziele dem Wunsch nach einer Bereicherung durch Musik im Weg steht (Ruf, 2014). Die Vermittlung musiktheoretischer Inhalte ereignet sich einerseits zwischen den zwei Polen *Spaß und Freude* sowie *notwendigem Wissen*, die zum einen affektiv-motivational, zum anderen kognitiv begründet sind und scheinbar wenig gemeinsame Schnittmenge aufweisen. Andererseits nehmen die curricularen Vorgaben und die Vergabe von Zensuren eine Metaposition ein, die ebenfalls auf die Ziele von Musikunterricht, beziehungsweise das Unterrichtshandeln von Musiklehrkräften einwirkt.

Vor diesem Hintergrund konnte auch Bieneck-Hempel (2009) im Rahmen ihrer empirischen Untersuchung zu Musiktheorie im Musikunterricht zeigen, dass sich Musiklehrkräfte hinsichtlich ihres Fachwissens zwar kompetent ausgebildet fühlen, es ihnen jedoch an fachdidaktischem Wissen sowie methodischen Strategien bezüglich der Vermittlung fehle. Ein Grund dafür könnte nach Richter (2002) in der Ausbildung von Musiklehrkräften liegen, der unter dem Begriff *Abbilddidaktik* kritisch zusammenfasst, dass „in Lehrplänen, Schulbüchern und in der Unterrichtspraxis, häufig in unreflektierter Gewöhnung, Auffassungen und Methoden einfach so übernommen, wie sie in den Hochschulen – mehr oder weniger gut begründet – betrieben werden“ (S. 46). Er bezieht sich dabei auf die Inhalte und Methoden, die im musiktheoretischen Teil der Ausbildung von Musiklehrkräften eingesetzt und in manchen Fällen dann von Lehrkräften „zu Zielen und Methoden des Unterrichts“ (Richter, 2002, S. 46) erhoben werden. Dabei wird deutlich, dass es an einem didaktischen Konzept für angehende Musiklehrkräfte hinsichtlich der Vermittlung musiktheoretischer Unterrichtsinhalte fehlt (siehe Richter, 2002). Die Entwicklung musikbezogener Strategien im Erklären und Instruieren könnte dahingehend einen Beitrag leisten, der für die Ausbildung für Lehrkräfte fachübergreifend – aber auch mit Blick auf den Musikunterricht – notwendig wäre (siehe Ehras et al., 2021).

Zusammenfassend kann festgestellt werden, dass die Frage, *warum* musiktheoretische Inhalte im Musikunterricht vermittelt werden sollten, mit den methodischen Herausforderungen der Vermittlung in Verbindung gebracht werden kann. Dabei stellen musiktheoretische Inhalte sowohl aus der Tradition als auch aus curricularer Sicht zentrale Elemente von Musikunterricht dar. Obwohl immer wieder Kritik daran geäußert wird, dass Musiktheorie im Musikunterricht vermittelt wird, richtet sich diese vermutlich eher danach, *wie* eine Vermittlung sowohl den Adressat*innen als auch dem Gegenstand von

Musik als klingendem Phänomen gerecht werden kann. Abschließend wird demnach zusammengefasst, wie musiktheoretische Inhalte vermittelt werden und welche Rolle dabei das Erklären einnimmt.

Vermittlung musiktheoretischer Inhalte

In der musikdidaktischen Fachliteratur wird der Wunsch geäußert, die Vermittlung musiktheoretischer Inhalte mit einem musikpraktischen Nachvollzug zu kombinieren (z. B. Schäfer-Lembeck, 2003). Wenn eine Lehrkraft also erklärt, sollte sie entweder akustische Repräsentationsformen in Form von Hörbeispielen präsentieren oder die Schüler*innen zum (gemeinsamen) Singen oder Musizieren anregen. Bereits erwähnt wurde allerdings auch, dass das Erklären einer Lehrkraft auf den ersten (oberflächlichen) Blick in Methodensammlungen und musikdidaktischer Fachliteratur nur eine sehr unbedeutende Rolle spielt. Im Folgenden soll das bis hierhin bereits erwähnte musikpädagogische Schrifttum noch einmal genauer hinsichtlich der Vermittlung musiktheoretischer Inhalte im Allgemeinen und der diesbezüglichen Rolle des Erklärens im Besonderen beleuchtet werden.

Buschendorff (2010) formuliert in seiner Methodensammlung unter anderem Herangehensweisen und Aufgabenstellungen einer „praktischen Musiktheorie“ (S. 31). Dabei legt der Autor neben verschiedenen Beispielen an Notationsübungen auch auf Aufgaben Wert, die in Kombination mit Hören (z. B. Hörübungen, bei denen die Richtung gespielter Intervalle zu bestimmen ist; Buschendorff, 2010, S. 48) oder „in Verbindung mit Keyboardspiel“ (Spielen von Tonleitern; S. 50) auszuführen sind. Zwar empfiehlt Buschendorff (2010) an manchen Stellen auch, etwas zu *erläutern* (z. B. „Kurze Erläuterung zum Tongeschlecht Moll“, S. 52; insgesamt siehe S. 41–67), allerdings wäre hier bezüglich der Bedeutung das Wort *Erklären* angemessener. Ähnliches kann bei Niessen (2010) beobachtet werden, die über eine Unterrichtsstunde resümiert, dass eine Lehrkraft polyphone Musik am Beispiel eines Kanons „erläutert“ (S. 69) und das ungefähr so sei, als würde man das Prinzip eines Tisches am Beispiel eines runden Tisches mit einem Bein in der Mitte „erklären“ (S. 69). Auch hier wird deutlich, dass Erklären und Erläutern als Synonym verwendet wird – das entspricht der Beobachtung Ehlichs (2009), der diesbezüglich für die Linguistik eine präzise Verwendung fordert, die auch für eine detaillierte Beschreibung des Unterrichtshandelns von (Musik-)Lehrkräften dienlich sein kann.

Wenn von Erklären die Rede ist, fallen vor allem zwei Dinge auf: Einerseits ist auf Erklären während des Musizierprozesses von Lehrkraft und Schüler*innen zu verzichten (z. B. Eberhard, 2010). Andererseits gibt es, wie bereits dargelegt, klare Vorstellungen darüber, dass Erklärungen erst dann zu geben sind, wenn Schüler*innen über ausreichend musikalische Erfahrung verfügen (Fuchs, 2015; Gebauer, 2013). Allerdings zeigt Lang (2010) in seiner Untersuchung zur Förderung von Improvisationsprozessen, dass eine kurze Erklärung⁶⁴ musiktheoretischer Inhalte im Voraus der Improvisation mehr Vorteile bringt als eine Improvisation, der keine musiktheoretische Vermittlung vorausgegangen

64 Auch Lang (2010) spricht an dieser Stelle von „*Erläuterung* [Hervorhebung vom Autor hinzugefügt] der harmonischen Struktur“ (S. 258).

ist. Während es also für die Reihenfolge einer solchen Kombination sehr viel häufiger den Wunsch nach *Handeln vor Wissen* gibt, liegen Beispiele vor, die das Prinzip *Wissen vor Handeln* verfolgen und dies mit mehreren Vorteilen versehen.

Darüber hinaus ist beim Blick in Methodensammlungen festzustellen, dass an manchen Stellen Instruktionen vorgenommen werden, an denen eigentlich Erklärungen zu erwarten gewesen wären: Buschendorff (2010) verweist beim „schriftlichen Konstruieren von Tonleitern“ (S. 51) auf mehrere Methoden, empfiehlt dabei jedoch ein Vorgehen, das nicht erklärt, sondern instruiert. Vorgegeben werden genaue Handlungsschritte, welche Töne wie zu notieren sind. Hinsichtlich der Abstände, die nach der Notation der Stammtöne einer D-Dur-Tonleiter zu notieren sind, wird folgende Instruktion gegeben: „Ist der Abstand zu klein, schreibe ein # vor die obere der beiden Noten. Ist er zu groß, notiere ein b“ (S. 51). Diesbezüglich ist zwar denkbar, dass im Unterricht weitere Erklärungen folgen können, allerdings wird ihre Notwendigkeit bei dieser Methodenempfehlung weder erwähnt noch auf sie hingewiesen. Ähnliche Vorgehensweisen finden sich auch bei der bereits erwähnten Einführung von Akkordumkehrungen (Nowak, 1999). Dabei wird aufgrund der Instruktionen der Eindruck erweckt, im Vordergrund steht das rezeptartige Bilden von Umkehrungen und weniger deren klangliche Charakteristika. Zwar wird dieser Vorgang bildlich gestützt, indem ein Dreiklang im Violinechlüssel nacheinander in Grundstellung, erster sowie zweiter Umkehrung und wieder Grundstellung angegeben wird, wobei Pfeile die angesprochene Oktavierung des Grundtons nachzeichnen; was diesbezüglich jedoch fehlt, sind Erklärungen innerhalb der Abbildung hinsichtlich des funktionalen Zusammenhangs zwischen dem tiefsten klingenden Ton, der das entscheidende Charakteristikum einer Grundstellung und ihrer Umkehrungen darstellt, und der Bezeichnung der Stellung (Nowak, 1999). Es wird vermutet, dass Handlungen auf Basis dieser Instruktionen zwar ausgeführt werden können, die zugrundeliegenden Wissenszusammenhänge jedoch nicht klar werden.

Hinsichtlich der *Dinge*, die bei der Vermittlung musiktheoretischer Inhalte zum Einsatz kommen, liegen unterschiedliche Vorstellungen vor, was am Beispiel der Klaviatur dargelegt wird: Nach Godau (2018) erweist sich in Aussagen von Musiklehrkräften die Nutzung einer Klaviatur als „typisches Musikunterrichts-Ding in ihrer Logik für die Vermittlung von Musiktheorie als unübertroffen stabil“ (S. 53) – beispielsweise dient die Klaviatur in Schulbüchern als Systematik zur Beschreibung von Intervallen und Tonleitern (z. B. Kemmelmeyer & Nykrin, 2004) oder wird in Methodensammlungen als Werkzeug zur Hilfe vorgeschlagen (z. B. Buschendorff, 2010). Einen anderen Ansatz verfolgt Beiderwieden (2008), der im Kontext der Vermittlung von Musiklehre Musiklehrkräften die klare Empfehlung gibt: „Schaffen Sie die Klaviatur ab“ (S. 103). Stattdessen schlägt der Autor vor, für die Vermittlung von Intervallen, Dreiklängen und Tonleitern einen diatonischen oder chromatischen Zirkel zu verwenden, die beide bildlich einem Treppenhäus gleichen, das aus der Vogelperspektive betrachtet wird (Beiderwieden, 2008). Die Verwendung dieser Zirkel verhindere nicht nur, Schüler*innen zu bevorzugen, die Klavier spielen können und für die eine Klaviatur etwas mehr oder weniger Selbstverständliches ist, sondern sie biete auch die Möglichkeit „eigentlich fast alles abbilden“ zu

können, „von der einfachen Notenlehre bis hin zu Symmetrien von Zwölftonreihen“ (Beiderwieden, 2008, S. 103).

Zusammenfassend ist festzuhalten, dass die Vermittlung musiktheoretischer Inhalte trotz aller Kritik ein wichtiges Ziel von Musikunterricht ist. Die Kenntnis über Musiktheorie geht nach Heß (2018b) „mit kognitiven Umgangsweisen einher und spiegelt ein Verständnis von Musik als geistig-kulturelles Phänomen wider“ (S. 79), womit auch die Realisation eines ästhetischen Weltzugangs eingelöst werden kann (siehe Baumert, 2002). Die Rolle des Erklärens wird bei der Vermittlung musiktheoretischer Inhalte zwar implizit sichtbar, jedoch in der musikdidaktischen Literatur nicht immer explizit benannt. Während einerseits häufiger von *Erläutern* gesprochen wird, sind andererseits in manchen Zusammenhängen eher Instruktionshandlungen zu beobachten, an deren Stelle eigentlich Erklärungen zu erwarten wären. Weiterhin erhalten die für den Musikunterricht typischen Dinge ein relevantes Gewicht hinsichtlich der Vermittlung von Musiktheorie. In der Gesamtschau liegt jedoch weiterhin nicht nur hinsichtlich des Erklärens musiktheoretischer Inhalte, sondern auch bezüglich der dabei relevanten Qualitätsaspekte ein Desiderat vor. Die vorliegende Arbeit möchte daher einen Beitrag im Sinne einer Grundlagenforschung über gutes Erklären im Musikunterricht leisten und mögliche Qualitätsmerkmale identifizieren.

3 Zielsetzung und Fragestellungen

3.1 Zielsetzung

In den theoretischen Überlegungen von Kapitel 2 wurde der Forschungsgegenstand dargestellt. Neben der begrifflichen Klärung, was die vorliegende Arbeit unter unterrichtlichem Erklären einer Musiklehrkraft versteht (siehe Kap. 2.1), wurde der Gegenstand in die bildungswissenschaftlichen Diskurslinien eingebettet (siehe Kap. 2.2). Dabei wurde deutlich, dass die Frage nach Erklärqualität eng mit der nach Unterrichtsqualität verweben ist und aufgrund der Fokussierung auf das Unterrichtshandeln einer Lehrkraft auch Schnittmengen mit der Diskussion um deren professionelle Kompetenz aufweist. Aufgrund des Desiderats innerhalb der musikpädagogischen Forschung wurde außerdem aufgezeigt, dass Charakteristika und Qualitätsmerkmale unterrichtlichen Erklärens, die im Kontext anderer Unterrichtsfächer untersucht wurden, für den Musikunterricht fruchtbar gemacht werden können (siehe Kap. 2.3). Deutlich gemacht wurde zudem, dass neben der Konturierung und Adaption generischer Merkmale für das Fach Musik auch spezifische Aspekte wie eine akustische Repräsentation eine wichtige Rolle spielen. Für einen ersten Zugang zum unterrichtlichen Erklären im Fach Musik wurden musikkundliche Themen als besonders *erklärbedürftige* Gegenstände gewählt (siehe Kap. 2.4).

In Bezug auf die Frage nach der Erfassung von Erklärqualität wurden in Kapitel 2.2 auch verschiedene Erhebungserfahren vorgestellt, die in der empirischen Bildungsforschung zum Einsatz kommen. Die vorliegende Arbeit setzt Beurteilungsverfahren zur Beschreibung von Erklärqualität ein. Beim Einsatz dieser Verfahren zeigt sich im Kontext der Forschung zu Unterrichtsqualität, dass ein umfassendes Gesamtbild nur durch das Zusammentragen unterschiedlicher Perspektiven gezeichnet werden kann (Kleickmann et al., 2019). Verschiedene Befunde betonen dabei, dass sich Beurteilungen verschiedener Statusgruppen wie Schüler*innen, Lehrkräfte und externe Beurteilende in manchen Aspekten überschneiden, in anderen aber auch unterscheiden können (z. B. Clausen, 2002; Fauth et al., 2014; für die Musikpädagogik z. B. Göllner, 2017). Kunter und Baumert (2006) nehmen für die Qualität von Unterricht eine perspektivenspezifische Validität an, wonach für verschiedene Gruppen unterschiedliche Aspekte einen guten Unterricht ausmachen (siehe z. B. auch Fauth et al., 2020). Bei der Frage, was gutes Erklären im Musikunterricht bedeutet, ist demnach auch von Interesse, ob eine perspektivenspezifische Validität vorliegt.

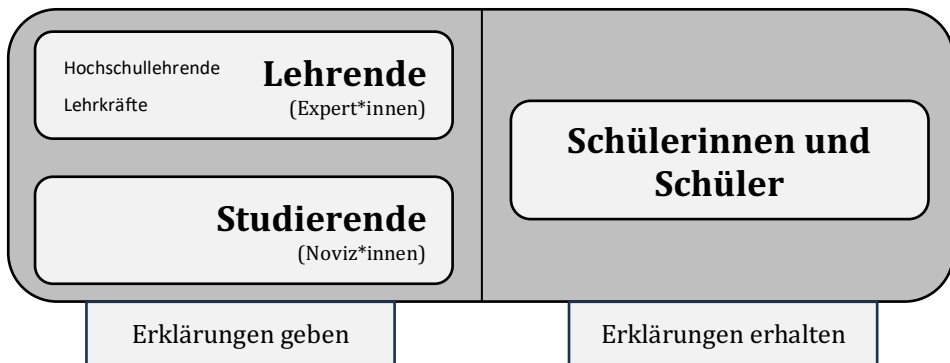
Um diese zu beantworten, wurde ein Onlinefragebogen entwickelt, der sowohl die Gesamtqualität einer Erklärung als auch Beurteilungen zu ausgewählten Qualitätsmerkmalen messen kann. Für eine spezifische Kontextualisierung der zu beurteilenden Erklärungssituation dienen kurze und unterrichtsnahe Erklärsequenzen in Form von Videovignetten, die als Stimulus in den Fragebogen implementiert werden. Das Erhebungsinstrument bildet damit einerseits Perspektiven verschiedener Statusgruppen auf Erklären ab. Andererseits ermöglichen offene und geschlossene Itemformate die Umsetzung explorativer sowie hypothesenprüfender Verfahrensweisen.

Für das erste zentrale Ziel, Unterschiede in der Bewertung der Erklärqualität zu untersuchen, werden zwei Aspekte fokussiert: Zum einen ist von Interesse, ob die Qualität unterrichtlichen Erklärens von verschiedenen Erklärmodi abhängt. Dabei steht der Einsatz optischer und akustische Repräsentationsformen im Zentrum. Zum anderen liegt ein Augenmerk darauf, inwieweit sich die Bewertungen von Erklärqualität zwischen folgenden drei⁶⁵ Statusgruppen unterscheiden:

- (1) Schüler*innen bewerten unterrichtliches Erklären aus der Perspektive von Adressat*innen, die eine Erklärung *erhalten*.⁶⁶
- (2) Studierende setzen sich innerhalb ihrer Ausbildung damit auseinander, Erklärungen zu *geben*.
- (3) Lehrende, die als Gruppe sowohl (Seminar-)Lehrkräfte als auch Fachdidaktiker*innen⁶⁷ umfassen, haben Erfahrung im *Geben* von Erklärungen.

Die beiden letzten Statusgruppen werden als *Erklärende* zusammengefasst und beinhalten Akteur*innen aus Schule und Lehrkräftebildung. Sie repräsentieren gegenüber den Schüler*innen die Sichtweise derer, die Erklärungen geben (siehe Abb. 3.1). Von Interesse ist sowohl, ob sich die Bewertungen von Erklärqualität aufgrund unterschiedlicher Expertise zwischen Studierenden und Lehrenden unterscheiden. Auch interessieren mögliche Unterschiede zwischen beiden Perspektiven: Schüler*innen auf der einen Seite und Studierende und Lehrende auf der anderen Seite.

Abbildung 3.1 Zusammensetzung der Statusgruppen



65 In einer anderen Veröffentlichung zur Studie FALKE-q-Mu erfolgt die Unterteilung in vier Statusgruppen aus Gründen der fächerübergreifenden Vergleichbarkeit (siehe Frei et al., i. Vorb.). Die Gruppe der Lehrenden wird dort nochmals in Lehrkräfte und Hochschullehrende (Didaktiker*innen) aufgeteilt.

66 Erklärungen von Schüler*innen an Klassenkamerad*innen sind kein Gegenstand dieser Untersuchung.

67 Gemeint sind Personen, die in der Hochschullehre an der Ausbildung von angehenden Lehrkräften beteiligt sind. Bei der Analyse der offenen Antworten wird zwischen Aussagen von Professor*innen und weiteren Hochschulmitarbeitenden (zusammengefasst als Hochschullehrende) unterschieden.

Das zweite zentrale Ziel besteht in der Identifizierung von Qualitätsmerkmalen unterrichtlichen Erklärens im Fach Musik. Grundlage dafür ist das theoretisch hergeleitete Modell guten Erklärens im Musikunterricht (Frei et al., 2022; siehe Kap. 2.4.2), das generische Merkmale von Erklären aufgreift und um fachspezifische ergänzt. Die empirische Untersuchung zu Relevanz und Wechselwirkung der ausgewählten Merkmale erfolgt getrennt nach Statusgruppen.

Ein drittes Ziel der vorliegenden Arbeit besteht darin, erste Hinweise auf den Einfluss des Vorwissens der Schüler*innen auf deren Bewertung der Erklärqualität zu erhalten. Der Grund dafür liegt in der Besonderheit des Fachs Musik, wonach vielfältige außerschulische Angebote zu großer Heterogenität zwischen Schüler*innen hinsichtlich musikalischer Vorerfahrungen führen (Ardila-Mantilla, 2018; Linn, 2017). Im Gegensatz zu den ersten beiden Zielen ist dieses jedoch nachrangig zu behandeln, da das Augenmerk dieser Studie auf Qualitätsmerkmalen unterrichtlichen Erklärens liegt und aus forschungsökonomischen Gründen keine umfassende Erhebung des Vorwissens möglich war. Ein kurzer Wissenstest (siehe Kap. 4.4.2) soll jedoch erste Anhaltspunkte zu dieser wichtigen Frage liefern, die dann für weitere Forschungsvorhaben als Grundlage dienen können.

3.2 Forschungsfragen und Hypothesen

Aus den Forschungszielen ergeben sich verschiedene Fragestellungen. Zunächst wird jedoch in einer vorbereitenden Analyse inferenzstatistisch überprüft, ob anhand der vorliegenden Daten die Aufteilung in drei Statusgruppen bestätigt werden kann (für die Analyse siehe Kap. 5.1; für die zugrunde liegende Forschungsfrage siehe Kap. 3.2.1). In den Formulierungen aller Forschungsfragen wird die Zugehörigkeit zu einer Statusgruppe als dreistufiger Faktor angenommen.

Die Forschungsfragen können entsprechend der beiden zentralen Ziele der vorliegenden Untersuchung in zwei Bereiche unterteilt werden: Zum einen ist zu klären, inwieweit sich die Bewertungen von Erklärqualität unter Berücksichtigung der Zugehörigkeit zur Statusgruppe sowie dem Einsatz verschiedener Erklärmodi unterscheiden (siehe Kap. 3.2.2). Zum anderen interessieren Zusammenhänge zwischen der Bewertung der Erklärqualität und Beurteilungen hinsichtlich verschiedener Qualitätsmerkmale unterrichtlichen Erklärens: Strukturiertheit, Adressatenorientierung, sprachliche Verständlichkeit, Sprech- und Körperausdruck und der Einsatz musikbezogener Repräsentationen⁶⁸ (siehe Kap. 3.2.3). Abschließend wird der Frage nachgegangen, inwieweit das Vorwissen der Schüler*innen Einfluss auf ihre Beurteilung von Erklärqualität nimmt (siehe Kap. 3.2.4).

68 Der Einsatz von Repräsentationen wird dabei auf zwei Ebenen untersucht: Zum einen als unabhängige Variable auf Videoebene (siehe dazu Kap. 4.3.1), zum anderen für jede Erklärung einzeln auf Skalenebene (siehe dazu Kap. 4.5.2).

3.2.1 Zusammensetzung der Statusgruppen

Innerhalb der Gruppe der Erklärenden können zwei Aufteilungen theoretisch begründet werden. Zum einen ergibt sich aus den unterschiedlichen Tätigkeitsfeldern eine dreiteilige Struktur: Studierende sind in der Ausbildungsphase, Lehrkräfte üben den Beruf aus und Didaktiker*innen, die als Hochschullehrende tätig sind, bilden Studierende in der ersten Phase der Lehrkräftebildung aus.⁶⁹ Zum anderen können die Profile von Lehrkräften und Didaktiker*innen aber auch verschwimmen: Didaktiker*innen waren als Musiklehrkräfte tätig, Lehrer*innen wiederum übernehmen in der Schule Aufgaben der Lehrkräftebildung (z. B. als Seminarlehrkräfte oder Praktikumslehrkräfte). Daher wird eine trennscharfe Abgrenzung beider Gruppen in Frage gestellt, die etwa mit Blick auf den Grad an Expertise zwischen Seminarlehrkräften und Professor*innen sehr spezifisch ausfallen würde und kein Gegenstand dieser Untersuchung ist. Aus den genannten Gründen erscheint eine gemeinsame Betrachtung von Lehrkräften und Didaktiker*innen als Gruppe von Lehrenden sinnvoll, die anhand der vorliegenden Daten inferenzstatistisch überprüft wird. Dabei werden signifikante Unterschiede zwischen Lehrkräften und Didaktiker*innen weder bei der Bewertung von Erklärqualität noch in den Beurteilungen zu ausgewählten Qualitätsmerkmalen erwartet. Forschungsfrage eins lautet daher:

F1) Unterscheiden sich die Bewertungen der Erklärqualität und die Beurteilungen zu Merkmalen guten Erklärens zwischen Lehrkräften und Didaktiker*innen signifikant?

Hypothese 1.1: Zwischen den Gruppen der Lehrkräfte und der Didaktiker*innen liegen keine signifikanten Unterschiede in den Globalbewertungen der Erklärqualität vor.

Hypothese 1.2: Zwischen den Gruppen der Lehrkräfte und der Didaktiker*innen liegen keine signifikanten Unterschiede in den Beurteilungen zu ausgewählten Qualitätsmerkmalen unterrichtlichen Erklärens vor.

3.2.2 Unterschiede in der Erklärqualität bezüglich Erklärmodus und Statusgruppe

Unterschiede in der Bewertung der Erklärqualität können verschiedene Gründe haben. Zwei mögliche Faktoren werden dabei herausgegriffen: Vor dem Hintergrund einer potenziellen perspektivenspezifischen Validität von Erklärqualität interessiert, inwieweit Unterschiede in der Bewertung anhand der Zugehörigkeit zu einer Statusgruppe erklärt werden können. Neben diesem strukturellen Aspekt erscheint für die Vermittlung musiktheoretischer Inhalte zudem ein inhaltlich-methodischer Faktor relevant: Beim Erklären

⁶⁹ Dieser Unterscheidung folgt das Projekt FALKE-q, aus dem heraus die Daten der vorliegenden Studie entstanden sind (Schilcher et al., 2021; Schilcher et al., i. Vorb.; für die Teilstudie des Fachs Musik, FALKE-q-Mu, siehe Frei et al., i. Vorb.).

kommen unterschiedliche Formen der Repräsentation zum Einsatz, die unterschiedlich kombiniert werden können und zu verschiedenen Erklärmodi führen. Während insbesondere Visualisierungen als optische Repräsentationsform fachübergreifend ein hoher Stellenwert beigemessen wird, ist für das Fach Musik unklar, welche Bedeutung im Verhältnis dazu der Einsatz akustischer Repräsentationen (z. B. Hör- oder Klangbeispiel) beim Erklären einnimmt. An dieser Stelle setzt Forschungsfrage zwei an, zu der keine Hypothesen formuliert werden:

F2) Wie unterscheiden sich die Bewertungen von Erklärqualität unter Berücksichtigung von Erklärmodus und Statusgruppe?

Aus dieser gehen zwei weitere Forschungsfragen hervor, die beide Faktoren gesondert in den Blick nehmen. Für den Faktor Statusgruppe interessiert vor allem, zwischen welchen Gruppen genau Unterschiede vorliegen. Aufgrund der Orientierung am Ansatz der Perspektiventriangulation – gemeint ist damit das gleichwertige Einbeziehen der Bewertungen aller drei Gruppen (siehe Kap. 4.1.1) – ist dabei nicht von primärem Interesse, wer die besseren oder schlechteren Noten gibt. Die Frage ist vielmehr, ob über alle drei Gruppen hinweg Einigkeit hinsichtlich der Qualität einer Erklärung besteht oder Unterschiede zwischen den Bewertungen der einzelnen Statusgruppen vorliegen:

F3) Wie unterscheiden sich die Bewertungen der Erklärqualität zwischen Schüler*innen, Studierenden und Lehrenden?

Im Zusammenhang damit ergeben sich zwei Teilfragen, für die jeweils Hypothesen formuliert werden können. Zum einen wird angenommen, dass die beiden Perspektiven – Erklärungen erhalten auf der einen (Schüler*innen) und Erklärungen geben auf der anderen Seite (Studierende und Lehrende) – für Unterschiede in der Bewertung von Erklärqualität verantwortlich sind. Zum anderen werden Unterschiede zwischen Studierenden und Lehrenden anhand des unterschiedlichen Grades an Expertise erwartet. Beide Hypothesen stehen in Verbindung mit Forschungsfrage drei:

(F3a) Liegen signifikante Unterschiede zwischen den Bewertungen der Schüler*innen auf der einen und Studierenden sowie Lehrenden auf der anderen Seite vor?

Hypothese 3.1: Die Bewertungen unterscheiden sich zwischen Schüler*innen und den Gruppen der Studierenden sowie Lehrenden signifikant.

(F3b) Unterscheiden sich die Bewertungen zwischen Studierenden und Lehrenden signifikant?

Hypothese 3.2: Die Bewertungen unterscheiden sich zwischen Studierenden und Lehrenden signifikant.

Beim Faktor Erklärmodus ist von Interesse, welche Bedeutung die beiden musikbezogenen Repräsentationsformen für die Bewertung der Erklärqualität haben. Im Gegensatz zum Faktor Statusgruppe ist demnach entscheidend, ob Unterschiede in der Bewertung mit den zum Einsatz kommenden Erklärmodi zusammenhängen. Forschungsfrage vier lautet daher:

F4) Wie unterscheiden sich die Bewertungen von Erklärqualität in Abhängigkeit verschiedener Erklärmodi?

Diese Frage wird in drei Teilfragen aufgegliedert. Zu erwarten ist, dass der gemeinsame Einsatz optischer und akustischer Repräsentationen aufgrund der vielschichtigen Zugänge das größte Potenzial an Erklärqualität aufweist.

(F4a) Werden Erklärungen, die beide Repräsentationsformen miteinander kombinieren, besser bewertet als solche, in denen nur eine Form der Repräsentation zum Einsatz kommt?

Hypothese 4.1: Erklärungen, die beide Repräsentationsformen miteinander kombinieren, werden besser bewertet als solche, in denen nur eine Form der Repräsentation zum Einsatz kommt.

Unklar ist jedoch, welche Bedeutung die beiden Repräsentationsformen für sich betrachtet auf die Qualität einer Erklärung haben. Denkbar ist zweierlei: Zum einen könnte die akustische Repräsentationsform besser bewertet werden und einen größeren Stellenwert einnehmen, weil nur dadurch dem Spezifikum von Musik als klingendes Phänomen Rechnung getragen werden kann (z. B. Geuen, 2018; Jank, 2013). Zum anderen könnte die optische Form der Repräsentation bevorzugt werden, da dies bei der Vermittlung musiktheoretischer Inhalte gängige Praxis zu sein scheint (z. B. Weidner, 2015). Aufgrund dieser beiden plausiblen Möglichkeiten wird zu folgender Forschungsfrage keine Hypothese formuliert:

(F4b) Werden Erklärungen besser bewertet, die ausschließlich eine optische Repräsentationsform verwenden, oder solche, die stattdessen eine akustische zum Einsatz bringen?

Zudem ist von Interesse, inwieweit der Einsatz von optischen und akustischen Repräsentationen in den offenen Antworten eine Rolle spielt, innerhalb derer die Bewertung zur Erklärqualität freiwillig begründet werden kann. Dabei geht es nicht um konkrete Hinweise zur tatsächlichen Umsetzung in der Erklärung. Vielmehr interessiert, ob in den Begründungen grundsätzliche Überzeugungen hinsichtlich des Einsatzes beider Repräsentationsformen beim Erklären im Musikunterricht herausgearbeitet werden können. An dieser Stelle setzt folgende explorative Frage an, zu der keine Hypothese vorliegt:

(F4c) Inwieweit spielen Überzeugungen zum Einsatz optischer oder akustischer Repräsentationen eine Rolle, wenn eine Bewertung von Erklärqualität begründet wird?

3.2.3 Zusammenhang zwischen Erklärqualität und Qualitätsmerkmalen

Neben der Frage, worin sich die Bewertungen von Erklärqualität unterscheiden, ist auch von Interesse, wie diese mit Beurteilungen zu ausgewählten Merkmalen unterrichtlichen Erklärens zusammenhängen. Welche Merkmale haben den größten Einfluss auf die Erklärqualität? Spielen andere Merkmale keine statistisch signifikante Rolle? Welche Gemeinsamkeiten und Unterschiede zeigen sich dabei zwischen den einzelnen Statusgruppen?

F5) Wie hängt die Bewertung der Erklärqualität mit Beurteilungen zu ausgewählten Qualitätsmerkmalen (Adressatenorientierung, Strukturiertheit, Sprachliche Verständlichkeit, Sprech- und Körperausdruck sowie Einsatz musikbezogener Repräsentationen) zusammen?

Die Auswahl der potenziellen Qualitätsmerkmale basiert auf Erkenntnissen fachübergreifender empirischer Forschung, weshalb signifikante Zusammenhänge mit der Erklärqualität für alle Kriterien erwartet werden. Hypothesen werden jedoch nur für das Merkmal Adressatenorientierung aufgestellt, da hierzu bereits mehrere empirische Befunde hinsichtlich der Wirksamkeit vorliegen (z. B. Nückles et al., 2005; Wittwer & Renkl, 2008).

(F5a) Besteht in allen drei Statusgruppen ein signifikanter Zusammenhang zwischen Erklärqualität und dem Merkmal Adressatenorientierung?

Hypothese 5.1: In der Gruppe der Schüler*innen liegt ein signifikanter Zusammenhang zwischen der Bewertung der Erklärqualität und der Beurteilung des Merkmals Adressatenorientierung vor.

Hypothese 5.2: In der Gruppe der Studierenden liegt ein signifikanter Zusammenhang zwischen der Bewertung der Erklärqualität und der Beurteilung des Merkmals Adressatenorientierung vor.

Hypothese 5.3: In der Gruppe der Lehrenden liegt ein signifikanter Zusammenhang zwischen der Bewertung der Erklärqualität und der Beurteilung des Merkmals Adressatenorientierung vor.

Die übrigen Merkmale unterliegen einer grundsätzlich explorativen Herangehensweise, da weder fundierte Annahmen hinsichtlich der Stärke der Zusammenhänge noch zu statusgruppenspezifischen Unterschieden getroffen werden können. Somit lautet die zweite Teilfrage:

(F5b) Wie unterscheidet sich die Bedeutung der einzelnen Qualitätsmerkmale zwischen den drei Statusgruppen?

Neben den hauptsächlich geschlossenen Itemformaten haben Studierende und Lehrende an verschiedenen Stellen des Fragebogens die Möglichkeit, in offenen Itemformaten zur Erklärung Stellung zu nehmen. Die getroffenen Aussagen liefern Hinweise auf subjektive

Theorien der Studienteilnehmer*innen und sollen einerseits auf relevante Äußerungen zu den ausgewählten Qualitätsmerkmalen unterrichtlichen Erklärens untersucht werden. Andererseits sollen dadurch auch blinde Flecken hinsichtlich weiterer Qualitätsmerkmale ausgeleuchtet werden, die in den geschlossenen Itemformaten nicht zum Tragen kommen werden. Aufgrund des explorativen Ansatzes werden keine Hypothesen in Zusammenhang mit Forschungsfrage sechs formuliert. Diese lautet:

F6) Welche Qualitätsmerkmale unterrichtlichen Erklärens haben in den subjektiven Theorien der Studienteilnehmer*innen Bedeutung?

(F6a) Welche Rolle spielen dabei die ausgewählten fünf Merkmale unterrichtlichen Erklärens?

(F6b) Welche weiteren Qualitätsmerkmale werden in den offenen Antworten genannt?

3.2.4 Zusammenhang zwischen Vorwissen der Schüler*innen und Erklärqualität

Ergänzend zu der Hypothese, dass das Merkmal Adressatenorientierung Bedeutung für die Qualität einer Erklärung hat, sollen in einer zusätzlichen Fragestellung erste Hinweise zur Frage gewonnen werden, wie das Niveau des Vorwissens der Schüler*innen mit deren Bewertung der Erklärqualität zusammenhängt. In anderen Fachkontexten wurde darauf hingewiesen, dass dieselbe Erklärung bei Schüler*innen zu unterschiedlichen Ergebnissen führen kann: Während niedrighschwellige Erklärungen insbesondere für leistungsschwächere Lernende hilfreich sind, können sie bei leistungsstarken Schüler*innen ihre Effektivität verlieren und sogar demotivierend wirken (siehe Findeisen, 2017). Übertragen auf das Fach Musik wird dementsprechend angenommen, dass das Vorwissen und die musikalische Vorerfahrung der Schüler*innen einen Einfluss darauf haben, wie sie die Qualität von Erklärungen bewerten. An dieser Stelle setzt Forschungsfrage sieben an:

F7) Wie hängt das Niveau des Vorwissens der Schüler*innen mit deren Bewertungen zur Erklärqualität zusammen?

Hypothese 7.1: Das Vorwissen der Schüler*innen hängt mit deren Bewertungen zur Erklärqualität zusammen.

4 Methode

Das folgende Kapitel erläutert das methodische Vorgehen der vorliegenden Studie. Neben einführenden forschungsmethodischen Überlegungen zur Triangulation von Perspektiven und Methoden werden das Forschungsdesign sowie die untersuchte Stichprobe vorgestellt. Da die Verwendung von Video-Vignetten als Stimulus eine spezifische Kontextualisierung der zu beurteilenden Erklärsituationen gewährleistet, werden die zugrunde liegenden Erklärgegenstände, die Vorgaben bei der Gestaltung und der Entstehungsprozess dargelegt. Die Videovignetten stellen das zentrale Element des Onlinefragebogens dar, der im Anschluss ebenso wie der ebenfalls integrierte Wissenstest für Schüler*innen vorgestellt wird. Es folgen Konzeptualisierung und Operationalisierung des Messgegenstands, an die sich eine Darstellung zusätzlich erhobener Kontrollvariablen anschließt. Nachdem der Ablauf der Erhebung erläutert wird, schließt das Kapitel mit der Vorstellung der verwendeten quantitativen und qualitativen Analyseverfahren.

4.1 Forschungsmethodische Überlegungen

Um dem komplexen Gegenstand unterrichtlichen Erklärens angemessen zu begegnen, die dargelegten Forschungsfragen zu beantworten und die Hypothesen zu prüfen (siehe Kap. 3.2), wird das Prinzip der Triangulation als vielversprechend erachtet. Es kommt in der vorliegenden Arbeit auf zwei Ebenen zum Einsatz: Einerseits werden Bewertungen zur Erklärqualität und Beurteilungen zu Qualitätsmerkmalen aus verschiedenen Perspektiven miteinander verglichen (siehe Kap. 4.1.1, Perspektivetriangulation), andererseits dienen unterschiedliche Methoden dazu, die Forschungsfragen zu beantworten (siehe Kap. 4.1.2, Methodentriangulation).

4.1.1 Perspektivetriangulation

In den theoretischen Ausführungen (siehe Kap. 2.2.1) wurden Beurteilungsverfahren, die verschiedene Perspektiven von Schüler*innen, Lehrkräften sowie externen Beurteilenden berücksichtigen, als bewährte Methode eingeführt, um die Qualität von Unterricht möglichst umfassend beschreiben zu können. Die vorliegende Arbeit orientiert sich an dieser Empfehlung, indem auch die Qualität des Erklärens aus unterschiedlichen Sichtweisen beurteilt werden soll. Dabei spielen zwei Aspekte eine wichtige Rolle.

Zum einen werden die drei Perspektiven der Schüler*innen, der Studierenden und der Lehrenden grundsätzlich als gleichwertig angesehen, sodass a priori keine Sichtweise stärker als eine andere gewichtet wird. Dabei können auf der einen Seite zwar insbesondere Lehrkräfte und Hochschullehrende aus der Gruppe der Lehrenden als Expert*innen für gutes Erklären angesehen werden, weshalb deren Urteil ein wichtiges Maß für die Qualität von Erklärungen darstellt. Auf der anderen Seite misst sich der Erfolg einer Erklärung jedoch vor allem daran, inwieweit Schüler*innen als Adressat*innen die

Erklärung positiv bewerten und deren Inhalte verstanden haben (siehe Kiel, 1999). Für die Gruppe der Studierenden wird angenommen, dass sie die Qualität einer Erklärung aus Sicht zukünftiger Lehrkräfte bewerten. Das spiegelt sich auch im Fragebogen wider, da Studierende und Lehrende dieselben Items beantworten und dabei die gleiche Perspektive einnehmen. Zwar verfügen Studierende in der Regel noch nicht über umfassende, mehrjährige Erfahrungen beim Erklären im Unterricht. Jedoch liegt die eigene Schulzeit noch nicht so lange zurück. Es wird demnach vermutet, dass Studierende eine dritte Sichtweise auf unterrichtliches Erklären einnehmen, weshalb auch ihre Perspektive in der Triangulation ebenfalls eine gleichwertige Rolle spielt.⁷⁰

Zum anderen werden für alle drei Statusgruppen dieselben Voraussetzungen für die Beurteilung geschaffen. Im Kontext der Forschung über Unterrichtsqualität stellt es ein häufig angewendetes Verfahren dar, dass Schüler*innen und Lehrkräfte Beurteilungen über ihren eigenen Unterricht abgeben, während externe Beurteilende diesen aus einer fremden Perspektive bewerten – Vor- und Nachteile der jeweiligen Perspektive diskutieren unter anderem Kleickmann et al. (2019). Um gerade der Kritik zu begegnen, dass aufgrund der subjektiven Involviertheit von Schüler*innen und Lehrenden in ihren eigenen Unterricht die Objektivität der Aussagen eingeschränkt ist, beurteilen in der vorliegenden Arbeit alle drei Gruppen Unterrichtssequenzen beziehungsweise Erklärungen einer fremden Lehrkraft. Mit Blick auf die Triangulation der drei Perspektiven kann dieser Aspekt zur Stärkung des Forschungsdesigns beitragen (siehe Kap. 4.2.1), da für alle Studienteilnehmer*innen gleichwertige Voraussetzungen für eine objektive Beurteilung geschaffen werden.

4.1.2 Methodentriangulation

Die Triangulation von Methoden zählt nach Flick (2011) zu den „Mixed Methodologies“, die primär eine „pragmatische Verknüpfung von qualitativer und quantitativer Forschung“ (S. 77) ermöglichen sollen. Döring und Bortz (2016) bekräftigen in diesem Zusammenhang ein neues Mixed-Methods-Paradigma, das auf Basis des philosophischen Pragmatismus die „Grundannahmen von quantitativem und qualitativem Paradigma nicht als unüberwindbare Grenze versteht“ (S. 74), sondern beide Forschungsstrategien als „legitim und gleichberechtigt“ (S. 73) anerkennt. Gläser-Zikuda et al. (2012) empfehlen ein solches Vorgehen unter anderem, um komplexen Forschungsgegenständen innerhalb der empirischen Unterrichtsforschung begegnen zu können. Es wird daher gewählt, um unterrichtliches Erklären zu untersuchen.

Die vorliegende Untersuchung verpflichtet sich dem Mixed-Methods-Paradigma und generiert dafür quantitative wie qualitative Daten. Als Erhebungsinstrument dient ein Fragebogen (siehe Kap. 4.4), der sowohl geschlossene als auch offene Itemformate⁷¹ enthält.

70 Dies gilt auch für die Auswertung der offenen Antworten, die sowohl von Studierenden als auch Lehrenden erhoben wurden (siehe Kap. 4.9.2). Beide Perspektiven werden auch dort grundsätzlich gleichwertig behandelt.

71 Auf das Interview als *klassische Form* der Erhebung qualitativer Daten wurde aus forschungsökonomischen Gründen verzichtet.

Durch die Kombination beider Datenformate werden verschiedene methodische Verfahren berücksichtigt (siehe Kap. 4.9), anhand derer sowohl explorative Forschungsfragen beantwortet als auch aufgestellte Hypothesen überprüft werden (siehe Kap 3.2). Den quantitativen Daten wird aus zwei forschungsökonomischen Gründen a priori jedoch ein stärkeres Gewicht beigemessen: Die qualitativen Daten wurden nur von Studierenden und Lehrenden erhoben – in einer Pilotierungsphase erwies sich die Form der Datenerhebung (offene Antwortfelder in einem Onlinefragebogen) für Schüler*innen aufgrund des hohen zeitlichen Aufwands als problematisch (Frei, 2019; siehe Kap. 4.7). Zudem war die Beantwortung der offenen Antworten für Studierende wie Lehrende fakultativ, was zu teils fehlenden, teils uneinheitlichen und damit insgesamt schwer zu vergleichenden Rückmeldungen in Umfang und Grad an Differenzierung führte. Mit Hilfe der quantitativen Daten wird außerdem der Großteil an Forschungsfragen beantwortet (F1, F2, F3, F4, F5, F7). Auch erfolgt die Prüfung von Hypothesen nur anhand quantitativer Daten. Die qualitativen Daten dienen vornehmlich explorativen Fragestellungen und übernehmen eine ergänzende Funktion. Sie sind insbesondere für die Beantwortung von Forschungsfrage sechs von großer Bedeutung.

4.2 Forschungsdesign und Stichprobe

Das Forschungsdesign orientiert sich am fachübergreifenden Rahmendesign der FALKE-q-Studie (Schilcher et al., i. Vorb.). So kann gewährleistet werden, die gewonnenen Daten einer vergleichenden Gesamtanalyse zu unterziehen (Lindl & Hilbert, i. Vorb.). Wie bereits dargelegt setzen sich die Statusgruppen in der vorliegenden Arbeit anders als bei der fachübergreifend vergleichenden Darstellung des Projekts FALKE-q für das Fach Musik (FALKE-q-Mu; siehe Frei et al., i. Vorb.) zusammen. Den Analysen liegen drei Statusgruppen zugrunde: Schüler*innen, Studierende und Lehrende.

4.2.1 Forschungsdesign

Die Anlage der Studie sieht vor, dass alle vier Statusgruppen dieselben sieben Videos zweimal beziehungsweise ausgewählte Abschnitte einzelner Videos ein drittes Mal beurteilen. Nach dem ersten Ansehen (Durchgang 1) bewerten die Studienteilnehmer*innen durch ein Globalurteil die Gesamtqualität der Erklärungen. Nach jedem Video geben sie eine holistische Beurteilung in Form einer Schulnote von eins bis sechs ab, die bei Bedarf mit einer Tendenz (+, –) versehen werden kann. Darüber hinaus besteht für die Gruppe der Erklärenden⁷² im Anschluss die Möglichkeit, diese Gesamtnote in einem offenen Antwortfeld zu begründen. An dieser Stelle des Fragebogens werden noch keine Items zu den Merkmalen unterrichtlichen Erklärens abgefragt, um die Aufmerksamkeit der Studienteilnehmer*innen bei den weiteren Videos nicht bereits auf einzelne Aspekte guten

72 Aus forschungsökonomischen Gründen wurde auf die Begründung der Globalnote bei der Gruppe der Schüler*innen verzichtet (siehe dazu Kap. 4.7).

Erklärens zu richten und ihre subjektive Wahrnehmung nicht zu beeinflussen. In einem zweiten Schritt (Durchgang 2) folgen dann drei beziehungsweise vier ausgewählte kurze Sequenzen der Erklärungen (ca. 30 Sekunden), die in Orientierung am Thin-Slices-Prinzip⁷³ (z. B. Babad, 2007) als Grundlage für die Bewertung des Kriteriums Sprech- und Körperausdruck der Lehrkraft dienen. Auch werden an dieser Stelle des Fragebogens Kontrollvariablen zur Persönlichkeitswirkung der Lehrkraft abgefragt. In einem dritten Schritt (Durchgang 3) sehen alle Studienteilnehmer*innen nochmals alle sieben Videos, wobei diesmal nach jedem Video Beurteilungen zu den in Kapitel 2 herausgearbeiteten Merkmalen unterrichtlichen Erklärens abgefragt werden. Abbildung 4.1 zeigt das Forschungsdesign.

Abbildung 4.1 Forschungsdesign

Onlinefragebogen (Beurteilung von 7 videografierten Erklärungen einer Musiklehrkraft)			
<i>Beurteilung derselben Videos</i>	7 Videos à ca. drei Minuten	3 bzw. 4 ausgewählte Videoausschnitte à ca. 30 Sekunden	7 Videos à ca. drei Minuten
Teilnehmer*innen (Statusgruppen):	Holistische Bewertung (Schulnote von 1 bis 6)	Kriterienbasierte Beurteilung (Ratingskala von 1 bis 6)	Kriterienbasierte Beurteilung (Ratingskala von 1 bis 6)
Schüler*innen Studierende Lehrende	Globalbewertung zur Qualität der Erklärungen	Items zu Sprech- und Körperausdruck sowie zur Persönlichkeitswirkung des Erklärenden	Items zu Einsatz musikbezogener Repräsentationen, Sprachliche Verständlichkeit, Strukturiertheit und Adressatenorientierung

4.2.2 Stichprobe

Die Stichprobe setzt sich aus Angehörigen der drei Statusgruppen zusammen und umfasst insgesamt $N = 271$ Studienteilnehmer*innen. Eingeschlossen sind dabei alle Personen, die mindestens zu einer Erklärung eine Bewertung abgegeben haben. Die Gesamtstichprobe verteilt sich über die Statusgruppen wie folgt: Die Gruppe der Adressat*innen einer Erklärung besteht aus $N_{SUS} = 148$ Schüler*innen⁷⁴, die Gruppe der *Erklärenden* umfasst

73 Babad (2007) empfiehlt bei der Untersuchung nonverbalen Verhaltens von Lehrkräften kurze, videografierte Sequenzen einer Unterrichtssituation. Dabei zeigt sich, dass Rater*innen ein Urteil über das nonverbale Verhalten schon auf Basis kurzer Szenen von weniger als einer Minute Dauer abgeben können.

74 Es lagen Daten von 149 Schüler*innen vor. Eine Person wurde aus dem Datensatz entfernt, da ein stereotypes Antwortmuster zu erkennen war. Dies zeigte sich zum einen in einer im Vergleich zur

dennach $N_{erkl} = 123$ Personen und setzt sich aus *Studierenden* und *Lehrenden* zusammen. Wegen der zum großen Teil auf persönlicher Kontaktaufnahme basierenden Rekrutierung liegt insgesamt eine nicht zufällige Gelegenheitsstichprobe vor.

Die Schüler*innen (kurz: SuS; $N_{sus} = 148$) kommen aus drei Gymnasialklassen zweier Gymnasien ($N_{sus.gym} = 78$) sowie aus drei Realschulklassen einer Realschule ($N_{sus.rs} = 70$) und wurden über Lehrkräfte sowie Partnerschulen aus dem Netzwerk der Universität Regensburg kontaktiert. Aufgrund von Datenschutzbestimmungen konnten bei der Gruppe der Schüler*innen das Geschlecht sowie weitere personenbezogene Merkmale nicht mit erhoben werden. Alle Schüler*innen besuchten zum Zeitpunkt der Befragung, die sich über den Zeitraum Februar und März 2018 erstreckte, die sechste Jahrgangsstufe.

Die zweite Statusgruppe bilden 61 Lehramtsstudierende (kurz: Stud; $N_{stud} = 61$), von denen 36 für das Lehramt an Gymnasien ($N_{stud.gym} = 36$) und 25 für das Lehramt an Realschulen ($N_{stud.rs} = 25$)⁷⁵ ausgebildet wurden. Über die verschiedenen Schulformen hinweg betrachtet waren 42 der Studierenden weiblich (69 %) und 19 (31 %) männlich. Sie befanden sich im Durchschnitt zwischen dem fünften und sechsten Fachsemester ($M = 5,70$), wobei die Semesterzahl eine hohe Streuung ($SD = 3,94$) aufweist. Die Erhebung der Studierenden erstreckte sich über einen Zeitraum von Oktober 2017 bis Dezember 2018, an der vor allem Studierende verschiedener Hochschulen und Universitäten aus Bayern⁷⁶ teilnahmen.

Die Gruppe der Lehrenden umfasst 62 Personen (kurz: Leh; $N_{leh} = 62$), von denen 28 weiblich (45 %) und 34 männlich (55 %) waren. Sie setzt sich einerseits aus Lehrkräften im Schuldienst ($N_{lk} = 41$) zusammen, die in den Schularten Gymnasium ($N_{lk.gym} = 22$) und Realschule ($N_{lk.rs} = 19$)⁷⁷ unterrichten und insgesamt 15 weibliche (37 %) und 26 männliche (63 %) Lehrkräfte umfassen. Der überwiegende Teil (30 Lehrkräfte, 73 %) hatte dabei Funktionen wie Seminar- oder Betreuungslehrkraft⁷⁸ beziehungsweise die Fachbetreuung inne, weshalb ein hoher Grad an Expertise angenommen wird. Die durchschnittliche Berufserfahrung entspricht etwa $M = 14,5$ Jahre, wobei auch hier eine breite Streuung in der Stichprobe vorliegt ($SD = 12,05$). Andererseits besteht die Gruppe der Lehrenden auch aus 21 Personen ($N_{did} = 21$), die in der Hochschullehre tätig und in die Ausbildung von angehenden Lehrkräften involviert sind und auf Unterricht eine Meta-Perspektive einnehmen. Ziel der deutschlandweiten Erhebung war, einen hohen Anteil an Professor*innen zu generieren. Von den 13 Didaktikerinnen (62 %) und acht Didaktikern (38 %) verfügen insgesamt 13 Personen über eine abgeschlossene Promotion und haben

Gesamtgruppe kürzeren Bearbeitungsdauer der Fragebögen, die ein Anklicken der Antwortkästchen ohne (konzentriertes) Lesen der Items vermuten lässt. Zum anderen konnte beobachtet werden, dass Items einer Skala immer mit denselben Extremwerten der Zustimmung beantwortet wurden, was in den Analysen vielfach zu Ausreißerwerten führte, die jeweils derselben Person zugeordnet werden konnten.

75 Zwei Teilnehmer*innen studierten Lehramt Grundschule, werden jedoch aufgrund eines ähnlichen Leistungsumfangs im Studium dieser Gruppe zugeordnet.

76 Zwei teilnehmende Studierende entstammen einer Universität außerhalb Bayerns.

77 Eine teilnehmende Lehrkraft unterrichtet an einer Grundschule. Sie wird aufgrund eines ähnlichen Leistungsumfangs im Studium der Gruppe der Realschullehrkräfte zugeordnet.

78 Eine Seminarlehrkraft ist in Bayern für die Ausbildung von Lehrkräften in der zweiten Phase zuständig. Eine Betreuungslehrkraft betreut Referendar*innen beim Einsatz außerhalb der Seminarschule.

teilweise eine Professur ($N_{did,prof} = 7$) inne. Daher kann auch für die Teilgruppe der Didaktiker*innen ein hoher Grad an Expertise angenommen werden. Während die Datenerhebung der Lehrkräfte im Rahmen einer Kooperationsveranstaltung für bayerische Seminar- und Praktikumslehrkräfte beider Schularten im Oktober 2017 startete und sich durch Hinzunahme weiterer Lehrkräfte bis Dezember 2018 erstreckte, erfolgte die Rekrutierung der Didaktiker*innen im selben Zeitraum vorwiegend über persönliche Kontaktaufnahme.

Tabelle 4.1 Überblick der Gesamtstichprobe

Statusgruppe	<i>N</i>	Geschlecht w/m/d	Schulart GS/MS/RS/GY	Jahrgangsstufe bzw. Erfahrung <i>M (SD)</i>	Funktion bzw. Akademischer Grad
Schüler*innen	148	---	---/---/70/78	Jgst.: 6	---
Studierende	61	42/19/---	2/---/23/36	FS.: 5,7 (3,94)	---
Lehrende	62	28/34/---	---	---	---
Lehrkräfte	41	15/26/---	1/---/18/22	BJ.: 14,47 (12,05)	12 BL/3 FB/15 SL
Didaktiker*innen	21	13/ 8/---	---/---/---/---	---	8 Prom./6 Dr./7 Prof.
Σ	271	70/53/---	3/---/111/136		

Bem.: GS: Grundschule, MS: Mittelschule, RS: Realschule, GY: Gymnasium; M: Mittelwert, SD: Standardabweichung; Jgst.: Jahrgangsstufe; FS: Fachsemester, BJ: Berufsjahre; BL: Betreuungslehrkraft, FB: Fachbetreuer, SL: Seminarlehrkraft; Prom.: Promovierende, Dr.: Doktor*innen, Prof.: Professor*innen.

4.3 Videovignetten

Die Videovignetten dienen als Stimulus zur Beurteilung der Erklärsituationen und spielen für die Entwicklung des Erhebungsinstruments eine zentrale Rolle. Dabei wird der Einsatz so genannter *scripted videos* (siehe Piwowar et al., 2018) aus zwei Gründen empfohlen: Durch die Videos, in denen Unterricht durch Schauspieler*innen nachgestellt wird, können datenschutzrechtliche Bestimmungen eingehalten werden, da kein offizieller Unterricht gefilmt werden muss. Andererseits liegt ein großer Vorteil darin, in der nachgestellten Situation den Fokus auf die Aspekte zu richten, die in der Untersuchung relevant sind. Um dennoch authentische Situationen zu generieren, fungiert eine Musiklehrkraft als Schauspieler in den Videovignetten. Dazu wird im Folgenden neben den inhaltlichen Beschreibungen der Erklärungen (siehe Kap. 4.3.1) auch angegeben, welche Gestaltungsvorgaben zur Erklärung an die Lehrkraft gerichtet wurden (siehe Kap. 4.3.2). Abschließend werden die eingesetzten Maßnahmen zur Qualitätssicherung erläutert (siehe Kap. 4.3.3).

4.3.1 Erklargegenstande

Die Auswahl der musiktheoretischen Inhalte, die fur die vorliegende Studie als Erklargegenstande dienen, orientiert sich an verschiedenen Aspekten. Erstens sind es Inhalte, die nach Weidner (2015) in nahezu jedem Curriculum allgemeinbildender Schulen zu finden sind. Zweitens sind die drei Gegenstande jeweils einem der drei Kompetenzbereiche Metrik, Rhythmik und tonaler Raum zuzuordnen, die dem Lehrwerk *Music Step by Step* zugrunde liegen und als drei Strukturelemente von Musik verstanden werden (Jank & Schmidt-Oberlander, 2010). Drittens wurden die ausgewahlten Themen von Seminarlehrkraften in Gesprachen auf die Frage nach besonders *erklarbedurftigen* musiktheoretischen Inhalten genannt.

Im Bereich *Rhythmik* wird das Verhaltnis der verschiedenen Notenwerte (kurz: NW) zwischen der Ganzen Note, der Halben Note und der Viertelnote auf zwei Weisen erklart. Dabei thematisiert die Lehrkraft das unterschiedliche Aussehen der Notensymbole und fuhrt am Ende den Begriff *Schlag* ein. Im zweiten Bereich *Metrik* wird erklart, inwieweit sich die Taktarten (kurz: TA) 4/4-Takt, 3/4-Takt und 2/4-Takt unterscheiden. Ausgehend von einem Puls in Viertelschlagen werden verschiedene Gruppierungen vorgenommen, mit denen unterschiedliche Betonungen innerhalb der Takte einhergehen. Anschließend wird an einem Fuball-Rhythmusbeispiel illustriert, wie derselbe Rhythmus (als Abfolge verschiedener Notenwerte) in verschiedenen Taktarten (mit metrischen Unterschieden in den Betonungen) klanglich realisiert wird. Im dritten Bereich *tonaler Raum* wird der Aufbau der Dur-Tonleiter (kurz: TL) anhand des Liedes *Dracula Rock* erklart. Dabei wird die Tonleiter als stufenweise systematisierte Darstellung des leitereigenen Tonvorrats einer Tonart abgebildet. Auf Basis dessen wird die spezifische Anordnung von Ganz- und Halbtonschritten unter Bezugnahme auf das Intervall der Sekunde erklart.

Über diese drei Themen hinweg werden die optischen wie akustischen Reprasentationsformen⁷⁹ systematisch variiert, wodurch diese beiden auf Videoebene als unabhangige Variable (UV; abgekurzt mit o = optisch; a = akustisch; oa = optisch und akustisch) implementiert werden. Um den jeweiligen Einfluss der eingesetzten Reprasentationsform untersuchen zu konnen, liegt fur jedes Thema eine Erklrung vor, die sowohl optische wie akustische Reprasentationsformen zum Einsatz bringt und damit als Vergleichswert fur die anderen Versionen angesehen werden kann (Video B: TL_oa; Video D: NW_oa; Video F: TA_oa). Dadurch ergibt sich bei drei Themen und drei moglichen Kombinationen von Reprasentationsformen eine Mindestanzahl von sieben Erklrungen, die in der vorliegenden Untersuchung aufgrund forschungskonomischer Grunde nicht überschritten wird. Tabelle 4.2 fasst die Inhalte der einzelnen Videovignetten zusammen.⁸⁰

79 Bezuglich verbaler Formen von Reprasentation (z. B. Nennung von Analogien; siehe Kap. 2.4.2) liegt keine systematische Variation vor. So wurde beispielsweise in Erklrung NW_o auf verbaler Ebene eine Analogie in der Erklrung eingesetzt, die das Verhaltnis zwischen einer Ganzen, einer Halben und einer Viertelnote auf der einen Seite mit der Aufteilung einer Stunde in eine halbe Stunde und eine Viertelstunde auf der anderen Seite in Beziehung setzt.

80 Die Transkripte der einzelnen Erklrungen sind in *Anhang A* zu finden.

Tabelle 4.2 Übersicht zu den Inhalten der Videovignetten

Thema Rep.-form	Notenwerte (NW)	Taktarten (TA)	Tonleiter (TL)
optisch (o)	Video A (NW_o) Verschiedene Notenwerte werden unter Verwendung optischer Repräsentationen erklärt.		Video E (TL_o) Die Dur-Tonleiter wird unter Verwendung optischer Repräsentationen erklärt.
akustisch (a)		Video C (TA_a) Unterschiedliche Taktarten werden unter Verwendung akustischer Repräsentationen erklärt.	Video G (TL_a) Die Dur-Tonleiter wird unter Verwendung akustischer Repräsentationen erklärt.
optisch und akustisch (oa)	Video D (NW_oa) Verschiedene Notenwerte werden unter Verwendung optischer und akustischer Repräsentationen erklärt.	Video F (TA_oa) Unterschiedliche Taktarten werden unter Verwendung optischer und akustischer Repräsentationen erklärt.	Video B (TL_oa) Die Dur-Tonleiter wird unter Verwendung optischer und akustischer Repräsentationen erklärt.

Um Reihenfolgeeffekte zu vermeiden, werden die Erklärungen in ihrer Reihenfolge variiert. Jeder Statusgruppe werden dabei drei Versionen vorgelegt. Dadurch kann davon ausgegangen werden, dass sich diese Effekte über alle Versionen hinweg tendenziell ausmitteln. Wie aus Tabelle 4.3 hervorgeht, wurde bei der Variation der Reihenfolge darauf geachtet, dass sich sowohl das erklärte Thema als auch die gewählte Form der Repräsentation abwechseln. Auch wird die Reihenfolge der Themen in allen Versionen beibehalten: Es beginnt jeweils mit einer Erklärung zu Notenwerten (Video A: NW_o; Video D: NW_oa), danach folgen eine Erklärung zur Tonleiter (Video B: TL_oa; Video E: TL_o; Video G: TL_a) und eine zu Taktarten (Video C: TA_a; Video F: TA_oa).

Tabelle 4.3 Reihenfolge der Videovignetten

Version	Video	Erklärung
1	A - B - C - D - E - F - G	NW_o - TL_oa - TA_a - NW_oa - TL_o - TA_oa - TL_a
2	D - E - F - G - A - B - C	NW_oa - TL_o - TA_oa - TL_a - NW_o - TL_oa - TA_a
3	D - G - F - E - A - B - C	NW_oa - TL_a - TA_oa - TL_o - NW_o - TL_oa - TA_a

Optische Repräsentationen erfolgen in den Erklärungen anhand der Darstellung von Notenbeispielen, aber auch durch grafische Visualisierungen: Farblichen Markierungen heben die Betonungen in einem Takt hervor (Video F: TA_oa), horizontale und unterschiedlich lang angeordnete Kästchen stellen die Dauer von Tönen dar (Video A: NW_o; Video D: NW_oa) oder die Abbildung einer Klaviatur illustriert sowohl die Abfolge von Ganz- und Halbtönen als auch deren Position in einer Tonleiter (Video B: TL_oa; Video E: TL_o). Akustische Repräsentationen werden in den Videos wie folgt umgesetzt: Notenwerte werden am Klavier vorgespielt (Video D: NW_oa), Ausschnitte eines zuvor

gesungenen Stückes dienen dazu, das Tonmaterial einer Tonleiter akustisch darzustellen (Video B: TL_oa; Video G: TL_a) und Rhythmen werden von der Lehrkraft vorgeklatscht oder deren Betonungen durch ein zusätzliches Aufstampfen ergänzt (Video C: TA_a; Video F: TA_oa).

4.3.2 Gestaltungsvorgaben

Ein erfahrener Musiklehrer aus dem Gymnasium stellte sich für die Erstellung der Videovignetten freiwillig zur Verfügung. Er wurde gebeten, zu den drei in Kapitel 4.3.1 dargestellten Inhalten kurze monologische Erklärungen zu konzipieren und die Inhalte für eine 6. Jahrgangsstufe zu erklären. Die Lehrkraft erhielt dabei kein vorgegebenes Skript, wie es im Kontext von *scripted videos* zu erwarten wäre. Dadurch sollte die Erstellung möglichst authentischer und unterrichtsnaher Erklärungen gewährleistet werden. Um jedoch sowohl Vergleichbarkeit zwischen den Videos zu erzielen als auch die systematische Variation im experimentellen Design (siehe Kap. 4.2.1) umzusetzen, sollte die Lehrkraft bei der Gestaltung Vorgaben berücksichtigen. Diese werden im Folgenden begründet.

Im Vorfeld wurde festgelegt, welche Inhalte die Lehrkraft unter Einsatz akustischer, welche unter Einsatz optischer und welche unter Kombination beider Repräsentationsformen erklären soll (siehe Tab. 4.2).⁸¹ Dadurch kann die systematische Variation der musikbezogenen Repräsentationsformen als unabhängige Variable realisiert werden. Um die verschiedenen Erklärungen zu einem Thema darüber hinaus vergleichbar zu halten, sollten übrige Parameter wie das Rhythmusbeispiel bei den verschiedenen Erklärungen zum Thema Taktarten oder die Klaviatur als Visualisierungshilfe von Ganz- und Halbtönen beim Thema Tonleiter, möglichst in allen Erklärungen eines Themas enthalten sein. Weitere für den Zweck der Vergleichbarkeit möglichst gleich gehaltene Rahmenbedingungen waren, dass die Erklärungen etwa drei Minuten lang dauern und monologisch präsentiert werden sollten. Die letzte Vorgabe verfolgt den Zweck, dass die erklärende Lehrkraft im Vordergrund steht und eine möglichst klare Situation geschaffen wird, in der das Erklären klar zu beurteilen ist und so Merkmale guten Erklärens identifizieren werden können. Aus diesem Grund wird auch auf Interaktion mit Schüler*innen in Form von Zwischenfragen oder gemeinsamen Musizierphasen verzichtet, obwohl damit ein wichtiges Charakteristikum unterrichtlichen Erklärens beziehungsweise ein bedeutsames Element von Musikunterricht ausgeblendet wird. Aus Gründen der Standardisierung wurde diese Limitation aber in Kauf genommen. Trotzdem treten in der Erklärungsituation einige ebenfalls freiwillig teilnehmende Schülerinnen⁸² als *Schauspielerinnen* in Erscheinung. So konnte der Lehrer

81 Das hatte auch zur Folge, dass der Lehrer an manchen Stellen anders erklären musste, als er das bei eigener Entscheidung in seinem Unterricht umgesetzt hätte. So äußerte er beispielsweise im Zusammenhang mit der Erklärung von Taktarten mit ausschließlich akustischen Repräsentationsformen (Video G: TA_a), dass er dieses Thema im Musikunterricht nie ohne optische Repräsentation erklären würde.

82 Dass alle Schülerinnen weiblich waren, hat weder einen besonderen Hintergrund noch ist es für die Erhebung relevant.

seine Erklärung adressieren, wodurch die ökologische Validität gesteigert werden sollte. Aus demselben Grund wurden die Erklärvideos zudem im gewohnten Musiksaal der Lehrkraft aufgenommen.

4.3.3 Entstehung und Qualitätssicherung

Der Lehrer wurde mit einer einzelnen Kamera in der Halbnahen⁸³ gefilmt. Eine weitestgehend gleichbleibende Kameraeinstellung gewährleistete, die Mimik und Gestik der Lehrkraft über mehrere Videos hinweg in vergleichbarer Weise wahrnehmen und beurteilen zu können. An einzelnen Stellen wurde allerdings zur besseren Sichtbarkeit in die hinter dem Lehrer projizierte Präsentation gezoomt. Um auch sprecherische Elemente möglichst gut abbilden zu können, erhielt die Lehrkraft ein Ansteckmikrofon; ein zusätzliches Mikrofon wurde in der Nähe des Klaviers angebracht, um die Klangbeispiele gut aufnehmen zu können. Zur Inhaltsvalidierung wurden die Videos Expert*innen aus der Lehrer*innenbildung vorgelegt und auf Grundlage deren Rückmeldungen modifiziert. So gab es insgesamt drei Aufnahmeeinheiten, aus denen mehrere Versionen zu jeder der sieben Erklärungen hervorgegangen sind. Die Auswahl der jeweils finalen Erklärung, die als Stimulus im Fragebogen eingesetzt wurde, erfolgte in Absprache mit der Lehrkraft sowie den bereits erwähnten Expert*innen.

4.4 Messinstrumente

Für die vorliegende Untersuchung wurden ein Onlinefragebogen sowie ein kurzer Wissenstest auf Basis der forschungsmethodischen Überlegungen neu konstruiert. Der Onlinefragebogen sollte in der Lage sein, verschiedene Beurteilungsperspektiven abzubilden und gleichermaßen explorative wie hypothesenprüfende Herangehensweisen zu ermöglichen. Der Kurz-Wissenstest diente dazu, erste Anhaltspunkte über den Zusammenhang zwischen Vorwissen der Schüler*innen und deren Bewertung der Erklärqualität zu generieren. Beide Instrumente werden im Folgenden vorgestellt.

4.4.1 Onlinefragebogen

Der eingesetzte und für die vorliegende Untersuchung neu entwickelte Fragebogen soll einerseits die standardisierte und objektive Durchführung einer Befragung gewährleisten und andererseits die Erhebung einer Stichprobe, bestehend aus Personen dreier Statusgruppen, unter möglichst geringem Zeitaufwand ermöglichen (Döring & Bortz, 2016, S. 414–416). Die Datenerhebung mittels Onlinefragebogen stellt dabei die direkte Übertragung digitaler Daten in eine Analysesoftware sicher, sodass Fehler bei der Datenübertragung zwischen Erhebungs- und Analyseinstrument minimiert werden können.

83 Die Halbnahe bezeichnet eine Einstellungsgröße auf die menschliche Figur von der Hüfte an aufwärts und ermöglicht gleichermaßen den Blick auf Mimik und Gestik (siehe z. B. Petrasch & Zinke, 2012).

Es gibt analog den beiden Perspektiven, Erklärungen geben und Erklärungen erhalten, auch zwei Versionen des Onlinefragebogens: Eine Version für die Perspektive der Schüler*innen und eine für die der Gruppe der Erklärenden. Dies hat an manchen Stellen zur Folge, dass Formulierungen der Items variieren oder einzelne Elemente wie die Kurz-Wissenstests nur in der Fassung für Schüler*innen zu bearbeiten sind. In dieser Version sind beispielsweise auch die Items hinsichtlich des Einsatzes musikbezogener Repräsentationsformen aus der Ich-Perspektive formuliert (siehe ausführlich Kap. 4.5). Die anderen zwei Statusgruppen (Studierende und Lehrende) sollen beurteilen, welche Relevanz *Schüler*innen einer sechsten Jahrgangsstufe* dem Einsatz im Durchschnitt beimessen. Ein weiterer Unterschied beider Versionen besteht darin, dass lediglich Studierende und Lehrende an verschiedenen Stellen im Fragebogen die Möglichkeit erhalten, ihre Bewertungen der Erklärqualität in Form offener Antworten zu begründen. Bei der Version des Fragebogens für Schüler*innen wurden die offenen Antworten aus forschungsökonomischen Gründen entfernt, um ausreichend Zeit für den zusätzlichen Wissenstest einzuräumen.

Der Fragebogen besteht aus drei Durchgängen. Zu Beginn von Durchgang 1 werden neben der Erstellung eines Teilnehmer-Codes demographische Daten abgefragt, wobei Studierende und Lehrende je nach Zugehörigkeit zur Statusgruppe unterschiedliche Informationen angeben.⁸⁴ Zunächst sehen alle Studienteilnehmer*innen nach einer kurzen Erläuterung der Rahmenbedingung zur Erstellung der Videos (siehe Kap. 4.3.2) und einer jeweils knappen Kontextualisierung der Erklärsituation (z. B. „Im Musikunterricht einer 6. Jahrgangsstufe werden gerade Notenwerte wiederholt.“) jede der sieben Videovignetten an. Dabei bewerten sie nach jeder Erklärsequenz zunächst die Gesamtqualität in Form eines Globalurteils mit Schulnoten von 1 bis 6 sowie einer bei Bedarf zu ergänzenden Tendenz (+, -); es können sich demnach Noten von 0,7 (1+) bis 6,3 (6-) ergeben. Die bereits erwähnten Kurz-Wissenstests für Schüler*innen, die jeweils vor und nach den ersten drei Videos zu bearbeiten sind, werden um die Frage nach ihrem bisherigen Wissen zu den erklärten Inhalten ergänzt. Sie erweitern die holistische Beurteilung an dieser Stelle der Erhebung ebenso wie die Möglichkeit für die Gruppe der Erklärenden, ihre Bewertung der Erklärung zu begründen. Zusätzlich geben die Schüler*innen eine Einschätzung darüber ab, ob sie die Erklärung im Video verstanden haben⁸⁵; auch die Teilnehmenden der Gruppe der Erklärenden schätzen nach der Erklärung ein, wie verständlich diese für Schüler*innen einer sechsten Jahrgangsstufe ist. An dieser Stelle der Erhebung werden bewusst noch keine Beurteilungen hinsichtlich der ausgewählten Qualitätsmerkmale unterrichtlichen Erklärens abgefragt, um die Aufmerksamkeit sowie die Beurteilungsgrundlage der Studienteilnehmer*innen nicht zu beeinflussen.

84 Folgende Variablen wurden gefragt: Geschlecht (Stud, Leh), Studienfach/Schulart (Stud), Fachsemester (Stud), Schulart (Lk), Berufsjahre (Lk), Funktionsstelle (Lk) und akademischer Grad (Did). Die Schulform der Schüler*innen konnte anhand der Fragebogenversion ermittelt werden, weshalb die Information vorliegt, ob sie in der Realschule oder im Gymnasium sind.

85 Bei dieser Frage wurde keine Zustimmungsskala verwendet. Vielmehr wurde nach dem Verstehen gefragt und die Antworten folgendermaßen skaliert: 1 = „sehr gut“, 2 = „gut“, 3 = „einigermaßen“, 4 = „nicht so gut“, 5 = „schlecht“, 6 = „sehr schlecht“.

Nachdem alle sieben Videos angesehen wurden, folgt Durchgang 2, in dem kurze Ausschnitte (Dauer jeweils ca. 30 Sekunden) aus drei beziehungsweise vier Erklärungen nochmals vorgespielt werden.⁸⁶ Nach einer knappen Kontextualisierung, worauf im Folgenden der Fokus zu richten ist, werden alle Studienteilnehmer*innen gebeten, den Sprech- und Körperausdruck der Lehrkraft in Form einer geschlossenen Itemskala zu beurteilen (siehe Kap. 4.5.1). Bei der Erhebung wurden dieser zweite Durchgang mit den kurzen Sequenzen immer direkt an den ersten Gesamtdurchgang der Videos angeschlossen. Dieses Vorgehen basiert auf der Annahme, dass sich diese Performanzaspekte an den 30-Sekunden-Ausschnitten (*Thin Slices*, siehe z. B. Babad, 2007) bewerten lassen; zudem wurde vermutet, dass auch die ausführlichen Erklärungen noch im Gedächtnis waren. Außerdem werden alle Statusgruppen gebeten, auf einer kurzen geschlossenen Itemskala Angaben über persönliche Eigenschaften der Lehrkraft zu machen, um somit auf deren Wirkung kontrollieren zu können (siehe Kap. 4.6.1).

Im abschließenden Durchgang 3 sehen alle Studienteilnehmer*innen nochmals alle sieben Videos. Nach jeder Erklärung beurteilen Studierende und Lehrende zunächst, ob Erklärsituationen aus dem Video so oder ähnlich im Musikunterricht vorkommen. Bei den Videovignetten, in denen entweder ohne den Einsatz einer optischen oder einer akustischen Repräsentationsform erklärt wird, sollen alle Studienteilnehmer*innen angeben, wie sehr sie sich die jeweils fehlende Repräsentationsform wünschen – die Schüler*innen geben diesen Wunsch aus ihrer eigenen Perspektive ab; die Erklärenden sollen für ihre Einschätzung Schüler*innen einer sechsten Jahrgangsstufe zum Maßstab nehmen (siehe Kap. 4.6.2). Zudem können Studierende und Lehrende dem Kollegen im Video Vorschläge zur Verbesserung der Erklärung in offenen Textfeldern unterbreiten.

Die anschließenden Fragen zu den ausgewählten Merkmalen unterrichtlichen Erklärens gliedern sich wie folgt und werden allen Studienteilnehmer*innen gestellt: Zunächst werden Einschätzungen jeweils zu den im Video vorkommenden musikbezogenen Repräsentationen abgefragt. Um dabei immer wieder an die getrennte Beurteilung der unterschiedlichen Repräsentationsformen (verbal, optisch, akustisch) zu erinnern, wird nach jedem Video zunächst die verbale Darstellung und im Anschluss daran die optischen und akustischen Repräsentationen bewertet (siehe Kap. 4.5.2, Tab. 4.5; MR_ver_1G, MR_opt_1G, MR_aku_1G). Nach diesem Block folgen die weiteren Items der Skala zu musikbezogenen Repräsentationen, die je nach Inhalt der Erklärung sowie den Repräsentationsformen unterschiedlich ausfallen (siehe Kap. 4.5.2). Bei all diesen Aspekten beurteilt die Gruppe der Erklärenden zusätzlich noch die konkrete Umsetzung der Lehrkraft (wieder in Form von Schulnoten) und hatte danach die Möglichkeit, dies in einem Textfeld zu begründen. Die Beurteilungen der weiteren Merkmale guten Erklärens erfolgt in nachstehender Reihenfolge: Sprachliche Verständlichkeit (siehe Kap. 4.5.3), Strukturiertheit (siehe Kap. 4.5.4) und Adressatenorientierung (siehe Kap. 4.5.5).

86 Da insgesamt sieben Videos vorlagen, wurden in Version 1 Sequenzen der ersten drei Videos (A, B und C) und in Version 2 und 3 Sequenzen der restlichen Videos (D, E, F und G) vorgespielt.

4.4.2 Wissenstest

Der Wissenstest wurde für die vorliegende Untersuchung neu entwickelt. Da die Frage nach dem Vorwissen nachrangig zu behandeln ist, wurde aus Zeitgründen kein umfangreiches Testinstrument konzipiert, das zudem nicht empirisch validiert ist. Die Bearbeitung des Kurz-Wissenstests erfolgte in Durchgang 1 des Schüler*innen-Fragebogens unter der Prämisse, dass die veranschlagte Zeit von insgesamt 40 Minuten nicht überschritten wird und die Erhebung innerhalb einer Schulstunde möglich ist (siehe Kap. 4.8). Für jedes Thema (Notenwerte, Taktarten und Tonleiter) wurden je drei Aufgaben entwickelt, die sowohl geschlossene als auch offene Antwortformate beinhalten (siehe *Anhang B*). Jede Aufgabe wird bei richtiger Lösung mit einem Punkt bewertet, sodass pro Thema drei und insgesamt neun Punkte maximal erreicht werden können. Als Beispiel eines Kurz-Wissenstests dient der zum Thema Notenwerte in Abbildung 4.2.

Beim Thema Notenwerte müssen die Schüler*innen zunächst aus fünf verschiedenen Notenwerten die Halbe Note auswählen, außerdem eine abgebildete Ganze Note ohne vorgegebene Antwortmöglichkeiten benennen und schließlich ebenfalls in einem offenen Antwortfeld angeben, wie viele Viertelnoten in eine Halbe Note passen.

Im Kontext der Dur-Tonleiter sollen die Schüler*innen aus vier vorgegebenen Möglichkeiten die C-Dur-Tonleiter auswählen, zudem benennen, wie viele Halbtonschritte es in einer Durtonleiter gibt, und dann noch aus vier Möglichkeiten die richtige Lösung auswählen, dass einer der Halbtonschritte zwischen dem zweiten und dritten Ton liegt.


Beim Thema Taktarten gestalten sich die Aufgaben folgendermaßen: Zuerst wird ein unvollständiger Dreiviertel-Takt gezeigt, in dem nur eine Halbe Note abgebildet ist. Die Schüler*innen können dazu aus fünf Antwortmöglichkeiten auswählen, wie der Takt vervollständigt werden muss. Ebenfalls muss aus fünf Antworten selektiert werden, dass in einem Dreiviertel-Takt der erste Schlag am stärksten betont ist. Schließlich wird gefragt, wie viele Halbe Noten in einen Zweiviertel-Takt passen.

Während die Schüler*innen alle drei Teile des kurzen Wissenstests beim ersten Videodurchgang direkt aufeinander folgend bearbeiten, wird jeder Test noch einmal nach dem jeweiligen Video gestellt, wobei dieser aus exakt denselben Aufgaben besteht. Einerseits kann so in der Tendenz ermittelt werden, wie hoch das Vorwissen der Schüler*innen zu den drei musiktheoretischen Inhalten ausfällt. Andererseits bietet die nochmalige Testung im Anschluss die Möglichkeit, erste Hinweise auf Veränderungen im Wissen der Schüler*innen zu erhalten.


Bei der Einordnung der Ergebnisse sind jedoch verschiedene Limitationen zu berücksichtigen: Aufgrund der Kürze des Tests beziehungsweise der Anzahl von lediglich drei Aufgaben sind geringe Reliabilitätswerte zu erwarten; zudem ist mit hohen Itemvarianzen im Vergleich zur Gesamtvarianz des Tests zu rechnen (Bühner, 2021, Kap. 4). Vor diesem Hintergrund sowie der geringen Anzahl an Schüler*innen verringert sich die Power des Tests. Daher muss damit gerechnet werden, dass potenzielle Unterschiede zwischen den beiden Testdurchläufen nicht signifikant werden (Bühner, 2021, Kap. 4). In diesem Fall erfolgt die Auswertung auf deskriptiver Ebene, auf deren Basis erste Hinweise hinsichtlich eines potenziellen Wissenszuwachses generiert werden.

Abbildung 4.2 Wissenstest Notenwerte

Klicke auf die Halbe Note!



Benenne den abgebildeten Notenwert!



Tippe die Antwort in das leere Feld.

weiß ich nicht

Wie viele Viertelnoten passen in eine Halbe Note?

Tippe die Antwort in das leere Feld.

weiß ich nicht

4.5 Operationalisierung des Messgegenstands

Basierend auf den theoretischen Ausführungen in Kapitel 2 wird im Folgenden dargelegt, wie die ausgewählten Merkmale guten Erklärens konzeptualisiert und operationalisiert sind. Die Zustimmung zu den einzelnen Items wurde jeweils auf einer sechsstufigen Skala gemessen: 1 = „stimme voll zu“, 2 = „stimme zu“, 3 = „stimme eher zu“, 4 = „stimme eher nicht zu“, 5 = „stimme nicht zu“, 6 = „stimme gar nicht zu“. In nahezu jeder Skala wird zudem ein Item integriert, das die Funktion eines Skalenglobalurteils einnimmt (z. B. „Die Erklärung war strukturiert“ für Strukturiertheit). Die Reihenfolge der im Folgenden dargestellten Merkmale guten Erklärens entspricht derjenigen im Fragebogen. Die sechsstufige Skala wurde gewählt, um die Beurteilung der Items analog zur Bewertung der Erklärqualität in Schulnoten zu gestalten. Dies soll insbesondere den Schüler*innen eine ihnen bekannte Form der Bewertung ermöglichen. Um die Triangulation der Perspektiven umsetzen zu können und die beiden Sichtweisen der Schüler*innen sowie der Gruppe der Erklärenden vergleichen zu können, lauten die Formulierungen der Items möglichst gleich.

4.5.1 Sprech- und Körperausdruck

Die Items zu *Sprech- und Körperausdruck*⁸⁷ (kurz: SK) lassen sich in paraverbale und extraverbale Aspekte einteilen (siehe Tab. 4.4)⁸⁸. Auf Basis der im Theorieteil eingeführten relevanten Faktoren (siehe Kap. 2.3.2) sollen sowohl Stimmklang, Artikulation, Sprechgeschwindigkeit, Prosodie und Sprechpausen (paraverbal beziehungsweise parasprachlich) als auch Körperausdruck und Proxemik⁸⁹ (extraverbale Aspekte) beurteilt werden. Die in Tabelle 4.4 genannten Items werden in allen drei Statusgruppen und in derselben Reihenfolge abgefragt – auch in den anderen im Projekt FALKE-q beteiligten Unterrichtsfächern diente diese Skala zur Beurteilung von *Sprech- und Körperausdruck* der jeweils erklärenden Lehrkraft (siehe Schilcher et al., i. Vorb.).

Tabelle 4.4 Operationalisierung *Sprech- und Körperausdruck*

Sprech- und Körperausdruck		
<i>Nr.</i>	<i>Itemformulierung</i>	<i>Perspektiven</i>
SK_1	Der Lehrer hatte eine angenehm klingende Stimme.	SL/SuS
SK_2	Der Lehrer hatte eine deutliche Aussprache.	SL/SuS
SK_3	Der Lehrer hatte eine angemessene Sprechgeschwindigkeit.	SL/SuS
SK_4	Der Lehrer hat Sprechpausen so gesetzt, dass ich gut folgen konnte.	SL/SuS
SK_5	Der Lehrer hatte eine abwechslungsreiche Sprechweise.	SL/SuS
SK_6	Der Lehrer hatte einen Körperausdruck, der die Erklärung unterstützt hat.	SL/SuS
SK_7	Der Lehrer hat sich den Schüler*innen zugewendet.	SL/SuS
SK_8G ⁹⁰	Die Art und Weise, wie der Lehrer spricht, gefällt mir.	SL/SuS

4.5.2 Einsatz musikbezogener Repräsentationen

Neben den Merkmalen guten Erklärens, die über alle Unterrichtsfächer des Projekts FALKE-q erhoben werden, liegt ein fachspezifischer Fokus auf dem *Einsatz musikbezogener Repräsentationen* (kurz: MR), denen beim Erklären musiktheoretischer Inhalte eine wichtige Rolle beigemessen wird. Die entwickelten Items werden in verbale, optische und akustische Formen der Repräsentation kategorisiert (siehe Tab. 4.5).

87 Ab dieser Stelle im Text werden alle Qualitätsmerkmale unterrichtlichen Erklärens (Adressatenorientierung, Strukturiertheit, Sprachliche Verständlichkeit, Sprech- und Körperausdruck sowie Einsatz musikbezogener Repräsentationen) sowie die Kontrollvariable Persönlichkeitswirkung im Fließtext als Eigennamen behandelt und kursiv gesetzt.

88 Sprachliche Gesichtspunkte werden in einem eigenen Merkmal Sprachliche Verständlichkeit abgefragt (siehe Kap. 4.5.3).

89 Proxemik beschreibt das Maß an interpersonaler Nähe und gegenseitiger Zugewandtheit (siehe z. B. Allhoff & Allhoff, 2010).

90 „G“ markiert das Skalen-Globalitem.

Aufgrund der Tatsache, dass die für die Erklärung ausgewählten Repräsentationen eng mit den jeweiligen Inhalten zusammenhängen und in manchen Videovignetten aufgrund der systematischen Variation (siehe Kap. 4.3.1) bestimmte Repräsentationsformen nicht vorkommen, gibt es keine einheitliche Skala, die sinnvoll über alle sieben Erklärungen erhoben werden kann.⁹¹ Vielmehr wird für jedes der drei Themen eine Kurzsкала⁹² (alle relevanten Items sind mit einem * versehen; siehe Tab. 4.5) mit jeweils vier Items konstruiert, um zwischen diesen drei Kurzsкаlen jeweils das gleiche Varianzspektrum annehmen zu können. Die jeweiligen Kurzsкаlen liegen den inferenzstatistischen Verfahren zugrunde. Die Skala erhebt somit den *Einsatz musikbezogener Repräsentationen* auf Basis dieser Kurzsкаlen in einer über die Videos hinweg vergleichenden Weise und ist an einigen Stellen um zusätzliche Items ergänzt. So können weitere spezifische inhaltliche Ausgestaltungen abgefragt werden; zudem ergibt sich dadurch für jede Erklärung eine Langskala (die Reliabilitätswerte sind für Kurz- und Langskala angegeben, siehe Kap 5.2.1). Aufgrund der singulären Erhebung bei einem einzelnen Video erlauben sie jedoch keine Vergleiche zwischen zwei Erklärungen desselben Inhalts (siehe Tab. 4.5, z. B. MR_opt_7, MR_aku_7). Neben den beiden Skalen für Notenwerte (nw_k⁹³) und Taktarten (ta_k), die für beide Videos dieselbe Kurzsкала verwenden, stellt das Thema Dur-Tonleiter aufgrund der insgesamt drei Erklärungen einen Sonderfall dar.

Gebildet werden hier zunächst zwei Kurzsкаlen, die für das Video mit rein optischer (tl.o_k) und für das Video mit rein akustischer Repräsentationsform (tl.a_k) entwickelt wurden und zudem jeweils in dem Video enthalten sind, das beide Repräsentationsformen kombiniert. Mit dem Ziel gleicher Varianzen pro Kurzsкала wurden aus diesen acht Items ebenfalls vier ausgewählt (tl.oa_k), die inhaltlich für die Skala am aussagekräftigsten erschienen. Neben den unterschiedlichen Item-Stämmen („Für mich war wichtig, ...“ für SuS bzw. „Für Schüler_innen einer 6. Jahrgangsstufe ist wichtig, ...“ für SL) werden die Itemformulierungen bei Bedarf an die beiden Perspektiven (SuS und SL) angepasst.

91 Beispielsweise macht es keinen Sinn, Items zu optischen Repräsentationsformen zu erfragen, wenn diese in der Erklärung nicht vorkommen.

92 Die Kurzsкаlen dienen zudem als Möglichkeit, fachübergreifende Vergleiche zu ermöglichen, da durch sie ein Konstrukt im Fach Musik vorliegt, das über die Videos und Themen hinweg eine fachspezifische Gesamterhebung dieses Qualitätsmerkmals ermöglicht.

93 Die Kurzschreibweise (nw_k_1) drückt aus, dass dieses Item in der Kurzsкала (k) zu beiden Erklärungen des Themas Notenwerte (nw) zum Einsatz kommt. Eventuelle Zahlen (siehe Tab. 4.5) geben die Position in der Skala an.

Tabelle 4.5 Operationalisierung Einsatz musikbezogener Repräsentationen

Einsatz musikbezogener Repräsentationen		
<i>Nr.</i>	<i>Itemformulierung</i> Für mich war wichtig, ... (SuS) Für Schüler_innen einer 6. Jahrgangsstufe ist wichtig, ... (SL)	<i>Perspektiven</i>
<u>Verbale Repräsentationsform</u>		
MR_ver_1G	... dass der Lehrer mit Worten erklärt hat.	SL/SuS ⁹⁴
MR_ver_2* (nw_k_1)	... dass der Lehrer auf Ähnlichkeiten mit der Mathematik hingewiesen hat.	SL/SuS
MR_ver_3* (tlo_k_1)	... dass der Lehrer die Töne auch „Bausteine“ genannt hat.	SL/SuS
MR_ver_4* (tla_k_1)	... dass der Lehrer das Beispiel mit den Legobausteinen genannt hat.	SL/SuS
MR_ver_5	... dass der Lehrer das Beispiel mit der Uhrzeit genannt hat.	SL/SuS
<u>Optische Repräsentationsform</u>		
MR_opt_1G* (nw_k_2) (tlo_k_2)	... dass der Lehrer eine Präsentation an der Tafel verwendet hat.	SL/SuS
MR_opt_2* (nw_k_3)	... dass der Lehrer die Notenwerte am Ende in unterschiedlich langen Kästchen angeordnet hat.	SL/SuS
MR_opt_3* (nw_k_4) (tlo_k_3) (tloa_k_1)	... dass sie an der Tafel sehen konnten, was der Lehrer mit Worten erklärt hat.	SL
	... dass ich an der Tafel sehen konnte, was der Lehrer mit Worten erklärt hat.	SuS
MR_opt_4* (tlo_k_4) (tloa_k_2)	... dass sie die Tonabstände durch die abgebildeten Klaviertasten sehen konnten.	SL
	... dass ich die Tonabstände durch die abgebildeten Klaviertasten sehen konnte.	SuS
MR_opt_5* (tla_k_2)	... dass sie an der Tafel sehen konnten, was der Lehrer am Klavier vorgespielt hat.	SL
	... dass ich an der Tafel sehen konnte, was der Lehrer am Klavier vorgespielt hat.	SuS
MR_opt_6* (ta_k_3)	... dass sie an der Tafel sehen konnten, was der Lehrer vorgeklatscht hat.	SL
	... dass ich an der Tafel sehen konnte, was der Lehrer vorgeklatscht hat.	SuS
MR_opt_7	... dass sie an der Tafel sehen konnten, wo Betonungen liegen.	SL
	... dass ich an der Tafel sehen konnte, wo Betonungen liegen.	SuS
<u>Akustische Repräsentationsform</u>		
MR_aku_1G* (tla_k_3) (ta_k_1)	... dass der Lehrer Hörbeispiele auf dem Klavier vorgespielt hat. (Video B, D und G)	SL/SuS
	... dass der Lehrer Hörbeispiele vorgeklatscht hat. (Video C und F)	
MR_aku_2* (ta_k_2)	... dass sie durch das Vorklatschen hören konnten, wo Betonungen sind.	SL

94 Die Abkürzung „SuS“ meint Schüler*innen; „SL“ steht für Studierende und Lehrende.

	... dass ich durch das Vorklatschen hören konnte, wo Betonungen sind.	SuS
MR_aku_3* (ta_k_3)	... dass sie die Unterschiede zwischen den Taktarten hören konnten.	SL
	... dass ich die Unterschiede zwischen den Taktarten hören konnte.	SuS
MR_aku_4* (ta_k_4)	... dass der Lehrer das, was er vorgeklatscht hat, auch gleich danach beschrieben hat.	SL/SuS
MR_aku_5* (tl.a_k_3)	... dass sie die Tonleiter am Klavier hören konnten.	SL
(tl.oa_k_3)	... dass ich die Tonleiter am Klavier hören konnte.	SuS
MR_aku_6* (tl.a_k_4)	... dass sie die unterschiedlichen Tonabstände hören konnten.	SL
(tl.oa_k_4)	... dass ich die unterschiedlichen Tonabstände hören konnte.	SuS
MR_aku_7	... dass der Lehrer zu den Tönen am Klavier mit dem Fuß gestampft hat.	SL/SuS
MR_aku_8	... dass sie durch das Vorspielen am Klavier hören konnten, wie lange die Töne dauern.	SL
	... dass ich durch das Vorspielen am Klavier hören konnte, wie lange die Töne dauern.	SuS
<u>Qualität musikbezogener Repräsentationsformen</u>		
MR_ver_1n	Bitte bewerten Sie mit einer Note, wie gut der Lehrer mit Worten erklärt hat. <i>(Skala: Note 1+ bis 6-)</i>	SL
Rf_ver	Hier können Sie Ihre Note gerne begründen. <i>(offene Antwort; optional)</i>	
MR_opt_1n	Bitte bewerten Sie mit einer Note, wie gut dem Lehrer die Präsentation an der Tafel gelungen ist. <i>(Skala: Note 1+ bis 6-)</i>	SL
Rf_opt	Hier können Sie Ihre Note gerne begründen. <i>(offene Antwort; optional)</i>	
MR_aku_1n	Bitte bewerten Sie mit einer Note, wie gut dem Lehrer die Hörbeispiele gelungen sind. <i>(Skala: Note 1+ bis 6-)</i>	SL
Rf_aku	Hier können Sie Ihre Note gerne begründen. <i>(offene Antwort; optional)</i>	

* = Item ist Bestandteil einer Kurzskaala (Zuordnung in Klammern, z. B. nw_k_1).

Bei allen Items des Konstrukts (MR_ver_2-5, MR_opt_2-7, MR_aku_2-8) wurde die Erklärenden-Gruppe (SL) um Folgendes gebeten: Bitte bewerten Sie mit einer Note, wie gut dem Lehrer dieser Aspekt gelungen ist. (Note 1+ bis 6-). Es liegen damit sowohl Einschätzungen über die Relevanz einzelner Merkmale für Schüler*innen als auch Beurteilungen über die Umsetzung durch die erklärende Lehrkraft für die Gruppen der Studierenden wie Lehrenden vor.

4.5.3 Sprachliche Verständlichkeit

Die Items zu *Sprachliche Verständlichkeit* (kurz: SV) greifen zentrale Aspekte der Verständlichkeits- und Mündlichkeitsforschung auf (siehe Tab. 4.6). Die Konzeptualisierung orientiert sich an verschiedenen sprachlichen Ebenen von Verständlichkeit. Neben der

generellen Verständlichkeit der Erklärung – in der Sprachwissenschaft verstanden als Text und somit hier an der Textebene orientiert – werden auch die Länge beziehungsweise die damit angenommene Komplexität der Sätze, die verwendeten Fachbegriffe beziehungsweise die komplexen Wörter sowie besondere Aufmerksamkeitsmarker wie *Das ist jetzt sehr wichtig* abgefragt. Somit liegt eine eher breite Erfassung des Konstrukts vor, wodurch eine Anschlussfähigkeit an sprachliche Analysen gewährleistet ist (siehe Thimmabrey & Lindl, i. Vorb.). Die allgemeine Textverständlichkeit – beziehungsweise in diesem Fall die generelle sprachliche Verständlichkeit der Erklärung – kann dabei als Globalskalenitem verstanden werden, das jedoch in der Skala als eines von vier Items aufgenommen ist. Alle im Folgenden genannten Items wurden in derselben Weise und Reihenfolge auch in den anderen im Projekt FALKE-q beteiligten Unterrichtsfächern abgefragt (siehe Schilcher et al., i. Vorb.). An einer Stelle (SV_3) ist die Itemformulierung der jeweiligen Perspektive angepasst.

Tabelle 4.6 Operationalisierung *Sprachliche Verständlichkeit*

Sprachliche Verständlichkeit		
<i>Nr.</i>	<i>Itemformulierung</i>	<i>Perspektiven</i>
SV_1	Der Lehrer hat sich gut verständlich ausgedrückt.	SL/SuS
SV_2	Manchmal hat der Lehrer extra gesagt, dass etwas sehr wichtig ist.	SL/SuS
SV_3	Bei manchen Wörtern wissen die Schüler_innen eventuell nicht, was sie bedeuten (n) ⁹⁵	SL
	Bei manchen Wörtern habe ich nicht gewusst, was sie bedeuten.	SuS
SV_4	Manche Sätze hat der Lehrer zu lang gemacht. (n)	SL/SuS

4.5.4 Strukturiertheit

Die Items zur *Strukturiertheit* (kurz: St) einer Erklärung umfassen die Klarheit über die Ziele einer Erklärung sowie den schrittweisen Aufbau, der beispielsweise einer inhaltlich-fachlichen, inhaltlich-chronologischen oder am Lernprozess der Schüler*innen orientierten Logik folgen kann (siehe Tab. 4.7). Neben diesen äußerlich-strukturellen Elementen werden zudem sowohl die Kohärenz als auch die Fokussierung auf wesentliche Inhalte als entscheidende Merkmale innerhalb einer Erklärung in die Konzeptualisierung aufgenommen. Alle im Folgenden genannten Items wurden in derselben Weise und Reihenfolge auch in den anderen an FALKE-q beteiligten Unterrichtsfächern abgefragt (siehe Schilcher et al., i. Vorb.) und sind für beide Perspektivgruppen gleich formuliert:

⁹⁵ „(n)“ indiziert die negative Polung des Items.

Tabelle 4.7 Operationalisierung *Strukturiertheit*

Strukturiertheit		
<i>Nr.</i>	<i>Itemformulierung</i>	<i>Perspektiven</i>
St_1	Es ist klar, was genau der Lehrer erklären wollte.	SL/SuS
St_2	Die Erklärung hatte einen roten Faden.	SL/SuS
St_3	Der Lehrer hat die Erklärung Schritt für Schritt aufgebaut.	SL/SuS
St_4	Der Lehrer hat zu viele Einzelheiten erklärt. (n)	SL/SuS
St_5G	Die Erklärung war gut strukturiert.	SL/SuS

4.5.5 Adressatenorientierung

Die Items zur *Adressatenorientierung* (kurz: Ad), also der optimalen Passung und Abstimmung der Erklärung auf relevante Kompetenzen der Schüler*innen, sind für die zwei einzelnen Perspektiven teilweise heterogen zusammengestellt (siehe Tab. 4.8). Aus beiden muss beantwortet werden, wie schwierig die Erklärung war (Ad_1), wie stark sie die Lebenswelt der Schüler*innen berücksichtigt (Ad_2) und inwieweit sie an ihr Vorwissen anknüpft (Ad_3), weshalb die Items jeweils unterschiedlich formuliert sind (SL und SuS).

Im Gegensatz dazu kann der Grad an didaktischer Reduktion aufgrund fehlender Expertise bei den Schüler*innen nur von der Gruppe der Erklärenden eingeschätzt werden (Ad_4); auch ist die Formulierung des globalen Skalenitems (Ad_6G) für die Schüler*innen nur schwer verständlich, weshalb auch dieses Item nur für die Gruppe der Erklärenden gestellt wird. Um diesbezüglich jedoch Vergleichbarkeit zwischen den beiden Perspektiven zu erlangen, wird die Frage nach der Passung (Ad_3) für beide Perspektiven gestellt – auch um einen Zusammenhang mit dem Globalitem (Ad_6G) der Skala herstellen zu können. Alle im Folgenden genannten Items wurden in derselben Weise und Reihenfolge auch in den anderen im Projekt FALKE-q beteiligten Unterrichtsfächern abgefragt (siehe Schilcher et al., i. Vorb.) und teils gleich (Ad_4), teils unter Berücksichtigung der beiden Perspektiven (z. B. Ad_2) formuliert:

Tabelle 4.8 Operationalisierung *Adressatenorientierung*

Adressatenorientierung		
<i>Nr.</i>	<i>Itemformulierung</i>	<i>Perspektiven</i>
Ad_1	Für Schüler_innen war die Erklärung zu schwierig.	SL
	Die Erklärung war für mich zu schwierig.	SuS
Ad_2	Die Erklärung berücksichtigte die Lebenswelt der Schüler_innen.	SL
	Die Inhalte der Erklärung hatten etwas mit meinem Leben zu tun.	SuS
Ad_3	Die Erklärung passte für Schüler_innen dieser Jahrgangsstufe.	SL
	Die Erklärung passte für mich.	SuS

Ad_4	Die Erklärung war angemessen didaktisch reduziert.	SL
Ad_5	Schüler_innen konnten der Erklärung mit ihrem Wissen gut folgen.	SL
	Ich konnte mit meinem Wissen gut folgen.	SuS
Ad_6G	Die Erklärung war adressatengerecht.	SL

4.6 Weitere erhobene Variablen

Neben diesen fünf Konstrukten wurden an einzelnen Stellen des Fragebogens noch Items erhoben, die in den späteren Analysen verschiedene Kontrollfunktionen übernehmen sollen. Dabei weisen die drei Items zur Persönlichkeit der erklärenden Lehrkraft (siehe Kap. 4.9) in den späteren Analysen – und über alle beteiligten Fächer der FALKE-q-Studie (siehe Schilcher et al., i. Vorb.) – eine hohe interne Konsistenz auf (siehe Lindl & Hilbert, i. Vorb.), weshalb sie im Folgenden als reflexives Konstrukt (Kap. 4.6.1) und weitere Kovariaten (Kap. 4.6.2) auf Itemebene dargestellt werden.

4.6.1 Persönlichkeitswirkung

Die *Persönlichkeitswirkung* (kurz: Pw) der Lehrkraft stellt kein Qualitätsmerkmal unterrichtlichen Erklärens dar. Die darin verwendeten drei Items (siehe Tab. 4.9) dienen vor allem dazu, in den Analysen – insbesondere im Kontext des Merkmals *Sprech- und Körperausdruck* – auf Sympathie- und Begeisterungseffekte (siehe z. B. Keller et al., 2016; Kunter et al., 2011) beziehungsweise die Natürlichkeit der Lehrkraft kontrollieren und die diesbezügliche Fehlervarianz abschätzen zu können. Die Daten wurden im Kontext von *Sprech- und Körperausdruck* am Ende der kurzen Videosequenzen erhoben und sind aus beiden Perspektiven gleich formuliert. Alle im Folgenden genannten Items wurden in derselben Weise und Reihenfolge auch in den anderen im Projekt FALKE-q beteiligten Unterrichtsfächern abgefragt (siehe Schilcher et al., i. Vorb.).

Tabelle 4.9 Operationalisierung *Persönlichkeitswirkung*

Persönlichkeitswirkung		
<i>Nr.</i>	<i>Itemformulierung</i>	<i>Perspektiven</i>
Pw_1	Die Lehrkraft wirkte in diesem Video auf mich sympathisch.	SL/SuS
Pw_2	Die Lehrkraft wirkte in diesem Video auf mich begeistert.	SL/SuS
Pw_3	Die Lehrkraft wirkte in diesem Video auf mich natürlich.	SL/SuS

4.6.2 Zusätzliche Kontrollvariablen

In den Fragebogen wurden noch weitere Kontrollvariablen (kurz: Kv) aufgenommen (siehe Tab. 4.10). Nach jeder Erklärung sollten die Schüler*innen in Durchgang 1

angeben, inwieweit sie das im Video Erklärte verstanden haben; dieselbe Einschätzung sollten Studierende und Lehrende über Schüler*innen einer sechsten Jahrgangsstufe treffen. Da kein ausführlicher Verständnistest in der Studie implementiert werden konnte, dient dieses Item (Kv_1) zur Ermittlung einer Tendenz bei den Schüler*innen – gleichwohl ist klar, dass diese Selbsteinschätzung keine genaue Aussage über das tatsächliche Verstehen liefert. Zusätzlich wurden die Schüler*innen, analog zum kurzen Vorwissentest nach den ersten drei Erklärungen in Durchgang 1, um die Einschätzung ihres jeweiligen Vorwissens zum gerade erklärten Thema gebeten (Kv_2).

In Durchgang 3 wurde jeweils nach den Videos, in denen keine optische oder akustische Repräsentationsform zum Einsatz gekommen ist, abgefragt, wie stark sich die Schüler*innen die jeweils fehlende Repräsentationsform wünschen würden (Kv_3). Zusätzlich gaben diese drei Statusgruppen an, inwieweit solche Erklärssituationen im Musikunterricht vorkommen (Kv_4). Die Antwortoptionen der Skalen sowie die dabei verwendeten Stufen sind bei jedem Item in Klammern angegeben. Fehlt diese Angabe, wurde eine sechsstufige Zustimmungsskala verwendet, wie sie bei den Merkmalen unterrichtlichen Erklärens zum Einsatz gekommen ist (siehe Kap. 4.5).

Tabelle 4.10 Operationalisierung *Zusätzliche Kontrollvariablen*

Zusätzliche Kontrollvariablen		
<i>Nr.</i>	<i>Itemformulierung</i>	<i>Perspektiven</i>
Kv_1	Was im Video erklärt wurde, verstehen Schüler_innen einer 6. Jahrgangsstufe ... (<i>sehr gut – gut – einigermaßen – nicht so gut – schlecht – sehr schlecht</i>)	SL
	Was im Video erklärt wurde, habe ich ... (<i>sehr gut verstanden – gut verstanden – einigermaßen verstanden – nicht so gut verstanden – schlecht verstanden – sehr schlecht verstanden</i>)	SuS
Kv_2	Was im Video erklärt wurde, ... (<i>wusste ich vorher schon – wusste ich teilweise schon – wusste ich vorher noch nicht</i>)	SuS
Kv_3_aku	Schüler_innen einer 6. Jahrgangsstufe würden sich wünschen, dass der Lehrer in der Erklärung auch Hörbeispiele verwendet. (Video A und E)	SL
	Ich hätte mir gewünscht, dass der Lehrer in der Erklärung auch Hörbeispiele verwendet. (Video A und E)	SuS
Kv_3_opt	Schüler_innen einer 6. Jahrgangsstufe würden sich wünschen, dass der Lehrer in der Erklärung auch die Tafel verwendet. (Video C und G)	SL
	Ich hätte mir gewünscht, dass der Lehrer in der Erklärung auch die Tafel verwendet. (Video C und G)	SuS
Kv_4	Erklärssituationen wie die im Video gibt es im Musikunterricht.	SL

4.6.3 Offene Itemformate

Aussagen in Form offener Antworten sind an drei Stellen des Fragebogens möglich: Im ersten Durchgang konnten die Studienteilnehmer*innen ihre Bewertung der Erklärqualität mit eigenen Worten begründen (Textfeld *Begründung*, kurz: Bg). Im dritten Durchgang finden sich zwei Stellen⁹⁶: Zum einen gibt es die Möglichkeit, einen Vorschlag zur Verbesserung der Erklärung zu unterbreiten (Textfeld *Verbesserung*, kurz: Vb). Zum anderen konnten offene Antworten zu den einzelnen Repräsentationsformen (optisch, akustisch und verbal) abgegeben werden, die in der Erklärung zum Einsatz gekommen sind (Textfeld *Repräsentationsformen*, kurz: Rf). Die Erhebung der offenen Antworten erfolgte computergestützt. Die entsprechenden Textfelder wurden in den Onlinefragebogen integriert. Tabelle 4.11 gibt eine Übersicht.

Tabelle 4.11 Offene Itemformate

Offene Itemformate		
Nr.	Itemformulierung	Perspektiven
1_Bg	Hier können Sie Ihre Note gerne kurz begründen. Bitte beschränken Sie sich auf maximal 3 Stichpunkte.	SL
2_Vb	Wenn Sie möchten: Wie kann der Kollege Ihrer Meinung nach die Erklärung verbessern? (optional) Bitte beschränken Sie sich auf maximal 3 Stichpunkte. <i>Alternativ:</i> Es gibt für mich nichts zu verbessern.	SL
3_Rf	Rf_ver Hier können Sie Ihre Note gerne kurz begründen. Bitte beschränken Sie sich auf maximal 3 Stichpunkte.	SL
	Rf_opt Hier können Sie Ihre Note gerne kurz begründen. Bitte beschränken Sie sich auf maximal 3 Stichpunkte.	
	Rf_aku Hier können Sie Ihre Note gerne kurz begründen. Bitte beschränken Sie sich auf maximal 3 Stichpunkte.	

4.7 Pilotierung des Erhebungsinstruments

Die Pilotierung des neu konstruierten Erhebungsinstruments erfolgte in mehreren Schritten und hinsichtlich der einzelnen Merkmalsskalen in unterschiedlicher Weise (siehe ausführlich Frei, 2019). Als erste Maßnahme des Pilotierungsprozesses wurden die theoretisch abgeleiteten und neu konstruierten Items auf Basis von Rückmeldungen einer Expert*innengruppe aus Lehrkräften und Didaktiker*innen selektiert beziehungsweise an einzelnen Stellen modifiziert. Die daraus hervorgegangene zweite Version des Fragebogens bearbeiteten Studierende anschließend unter Zuhilfenahme der *Probing*-Methode. Die auf verschiedenen Nachfragetechniken beruhende Methode gehört ebenso wie die

⁹⁶ In Durchgang 2 (Angaben zu den Variablen *Sprech- und Körperausdruck* und *Persönlichkeitswirkung*) gab es keine Möglichkeit für offene Antworten.

Think-Aloud-Methode zur Familie der kognitiven Interviews (siehe Prüfer & Rexroth, 2005). Letztgenannte wurde in einer frühen Phase der Pilotierung angewandt, erwies sich jedoch nicht als gewinnbringend, weshalb die Pilotierung ausschließlich mit der *Probing*-Methode durchgeführt wurde (Frei, 2019).

Auf Basis der Erkenntnisse, die durch die Pilotierung mit den Studierenden gewonnen wurden, resultierte eine dritte Version des Fragebogens. Diese wurde abschließend mit Schüler*innen aus Gymnasium sowie Realschule pilotiert, anhand deren Rückmeldungen letzte Modifikationen vorgenommen wurden, sodass daraus die vierte und endgültige Version des Fragebogens entstanden ist. In allen beschriebenen Phasen lag der Schwerpunkt der Pilotierung auf den Items zum *Einsatz musikbezogener Repräsentationen*.

Die Items der anderen fachübergreifend erhobenen Qualitätsmerkmale wurden von allen in FALKE-q beteiligten Fächern in einem gemeinsamen Vorgehen aller Projektbeteiligten pilotiert. Der Frage, wie die fachübergreifenden Merkmale *Sprachliche Verständlichkeit*, *Sprech- und Körperausdruck*, *Strukturiertheit* und *Adressatenorientierung* zu konzeptualisieren und zu operationalisieren sind, waren zudem mehrere Diskussionsrunden zwischen Mitarbeiter*innen und den betreuenden Professor*innen aller 14 Disziplinen des Projekts FALKE-q vorausgegangen. Die zentralen Erkenntnisse aus der Pilotierung der Skala zum Merkmal *Einsatz musikbezogener Repräsentationen* sind an anderer Stelle ausführlich zusammengefasst (Frei, 2019), sollen hier jedoch knapp dargelegt werden.

Erstens zeigte sich in der Pilotierung, dass die offenen Antworten der Schüler*innen zur Begründung des Globalurteils auf der einen Seite viel Zeit zur Beantwortung in Anspruch nahmen, die Ergebnisse auf der anderen Seite für das Erkenntnisinteresse aber oft nicht dienlich waren. Da zudem zu wenig Zeit für den Kurz-Wissenstests blieb, wurden die offenen Antworten für Schüler*innen entfernt. Zweitens dienten konkrete Nachfragen hinsichtlich einzelner Wortbedeutungen dazu, das jeweilige Verstehen der Studienteilnehmer*innen mit dem ursprünglich intendierten Sinn abzugleichen (Frei, 2019). Drittens konnte die Pilotierung mit Studierenden aufzeigen, dass der in dieser zweiten Version des Fragebogens durchgehend verwendete Zusatz *um die Erklärung zu verstehen* zu Problemen bei der Beantwortung führte. Mit der Intention, eine Zielperspektive der Erklärung auszudrücken, wurde angemerkt, dass das Item (*Das Beispiel mit der Uhrzeit war für mich wichtig, um die Erklärung zu verstehen*) nicht mehr zu beantworten ist, wenn der Inhalt eben nicht verstanden wurde (Frei, 2019). Viertens zeigte sich durch die Antworten zu den Nachfragen auch, dass teilweise bewertet wurde, *ob* eine Präsentation an der Tafel verwendet wurde (Relevanz des Vorkommens), und teilweise, *wie* die Präsentation umgesetzt wurde (Bewertung der Umsetzung); somit mischten sich Einschätzungen zu einem Aspekt des Erklärens (*Das Beispiel mit der Uhrzeit war für mich wichtig*) mit der bewertenden Selbsteinschätzung (*um die Erklärung zu verstehen*). Notwendig wurde daher eine Anpassung der Items, weshalb in der endgültigen Version sowohl nach dem Vorkommen als auch nach der Umsetzung gefragt wurde (Frei, 2019; siehe z. B. Kap. 4.5.2, Tab 4.5, MR_opt_1G und MR_opt_1n).

Im Gegensatz dazu zeigte sich fünftens bei den Schüler*innen, dass sie bei dieser zweigliedrigen Fragestrategie keine Unterscheidung zwischen beiden Items (Vorkommen

und Umsetzung) trafen, sondern ausschließlich Angaben dazu machen konnten, wie wichtig für sie das Vorkommen einer bestimmten Repräsentationsform war (Frei, 2019). Daher wurde die Frage nach der Umsetzung bei Schüler*innen nicht gestellt. Sechstens wurden weitere Items wie *Eine Erklärung wie die im Video wünsche ich mir für den Musikunterricht* aus dem Fragebogen aufgrund potenzieller Missverständnisse ausgeschlossen: Die Schüler*innen gaben eher noch einmal eine Gesamtbewertung ab, was dadurch ersichtlich wurde, dass sie nochmals erfragt haben, welche Note sie am Anfang für das Globalurteil abgegeben haben (Frei, 2019). Während die Intention des Items war herauszufinden, ob Schüler*innen solche monologischen Erklärungen von einer Lehrkraft ablehnen, könnte der Eindruck einer zu erteilenden Gesamtbewertung auch wegen der Platzierung des Items am Ende des Fragebogens erweckt worden sein (Frei, 2019).

Der auf Basis dieser ausführlichen Pilotierung entstandene Fragebogen wurde in beiden Versionen als Onlinefragebogen mittels des Survey-Tools Unipark⁹⁷ erstellt, um eine einfache und gerade in Verbindung mit dem Abspielen der Videos zweckmäßige Erhebung zu garantieren. Deren Durchführung wird im Folgenden erläutert.

4.8 Durchführung der Erhebung

Die insgesamt drei Durchgänge der Erklärvideos (siehe Kap. 4.4) wurden an zwei Zeitpunkten erhoben, weshalb die Fragebögen in zwei Teile geteilt wurden. Der erste Teil des Fragebogens umfasst sowohl Durchgang 1 als auch Durchgang 2 und hat bei der Version für Schüler*innen auch den Wissenstest integriert. Der zweite Teil des Fragebogens umfasst demnach für alle Studienteilnehmer*innen in Durchgang 3.

Für die Schüler*innen fand die Erhebung in den jeweiligen Computerräumen der beteiligten Gymnasien sowie der Realschule aus Regensburg statt. So konnten alle die Bearbeitung an einem eigenen Computer und während der Unterrichtszeit erledigen. Um einen reibungslosen Ablauf zu gewährleisten, waren neben der Lehrkraft dabei auch zwei Projektmitarbeitende anwesend, die nach einer kurzen Einführung und Erläuterung des Ablaufs auch für Verständnisfragen während der Bearbeitung des Fragebogens zur Verfügung standen (siehe *Anhang C2*). Diese dauerte für beide Teile des Fragebogens etwa 40 Minuten.

Die Studienteilnehmer*innen aus der Gruppe der Erklärenden nahmen meist an ihren eigenen Endgeräten in ihrer Freizeit teil. Als Ausnahme sind einige Studierende zu nennen, die beide Teile des Fragebogens im Rahmen von universitären Lehrveranstaltungen in zwei verschiedenen Sitzungen bearbeiteten, sowie Seminar- und Praktikumslehrkräfte, die innerhalb eines vom Projekt FALKE-q organisierten Kooperationstages zwischen erster und zweiter Phase der Lehrer*innenbildung an der Erhebung teilgenommen haben. Während der erste Teil des Fragebogens im Vorfeld der Veranstaltung zuhause ausgefüllt wurde, bearbeiteten alle teilnehmenden Lehrkräfte den zweiten Teil an diesem Tag vor Ort.

⁹⁷ Siehe <https://www.questback.com/de/online-befragungstool/>.

Alle anderen Personen dieser drei Gruppen (Studierende, Lehrkräfte, Didaktiker*innen) erledigten die beiden Teile nacheinander an eigenen Endgeräten. Um den Organisationsaufwand gering zu halten, wurden bereits beide Links zu den Fragebogenteilen in einer Mail versandt. Anhand der Bearbeitungszeitpunkte kann ausgeschlossen werden, dass beide Teile in falscher Reihenfolge bearbeitet wurden.

4.9 Analyseverfahren

Das abschließende Teilkapitel stellt dar, mit welchen Analyseverfahren die aufgeworfenen Fragestellungen bearbeitet beziehungsweise die Hypothesen geprüft werden. Dabei erfolgt eine getrennte Beschreibung der quantitativen und qualitativen Analyseverfahren.

4.9.1 Statistische Analyse

Stichprobe⁹⁸

Insgesamt haben 271 Studienteilnehmer*innen an der Erhebung teilgenommen. Da diese in zwei Teilen erfolgte, ergeben sich für beide Abschnitte unterschiedliche Fallzahlen. Für den ersten Teil des Fragebogens lagen insgesamt 268 Fälle vor – drei Schüler*innen nahmen nur am zweiten Messzeitpunkt, jedoch nicht am ersten teil. Kriterium für diese Zählung war, dass mindestens das erste Video mit einer Globalbewertung in Durchgang 1 beurteilt wurde. Von diesen liegen 263 komplett bearbeitete Fälle des ersten Fragebogenteils vor, die sich auf 145 Schüler*innen, 59 Studierende und 59 Lehrende aufteilen – an unterschiedlichen Stellen des ersten Abschnitts haben zwei Studierende und drei Lehrende den Fragebogen abgebrochen.

Analog der Zählung, dass zu mindestens einem Video alle Daten vorliegen, haben den zweiten Teil des Fragebogens insgesamt 254 Personen begonnen, davon 136 Schüler*innen, 59 Studierende und 59 Lehrende. Abgebrochen haben drei Studierende und drei Lehrende, sodass 248 komplette Fälle von beiden Fragebogenteilen vorliegen.

Daten zu allen Items beider Fragebogenteile liegen von 133 Schüler*innen, 56 Studierenden und 56 Lehrenden vor. Diese 245 vollständigen Fälle entsprechen gegenüber den insgesamt vorliegenden 271 Fällen einem Anteil von 90,06 %, was zu einer Drop-outrate von 9,6 % führt. Dieser Wert liegt nach Schnell et al. (2018) im üblichen Bereich sozialwissenschaftlicher Forschung. Der Umgang mit fehlenden Werten ist für die beiden Hauptanalysen unterschiedlich zu bewerten und wird an entsprechender Stelle erläutert.

⁹⁸ Da die Angabe der Stichprobe für die statistischen, nicht jedoch für die inhaltsanalytischen Auswertungen (siehe Kap. 4.9.2) relevant sind, werden sie aus Gründen der eindeutigen Zuordnung an dieser Stelle – und nicht wie üblich zu Beginn des Ergebnisteils – berichtet.

Die quantitativen Daten wurden mit der Statistiksoftware *R* (R Core Team, 2021) analysiert. Dabei erfolgte der erste Schritt der Datenaufbereitung bereits innerhalb der Onlinefragebogen-Software *Unipark*, indem der Datensatz bereits als numerische Ausgabe und mit Variablenbeschriftung exportiert wurde. In *R* wurden dann alle negativ formulierten Items⁹⁹ (z. B. Tab 4.7, St_4) recodiert und deren Zahlenwerte entsprechend der Polung der anderen Items in der jeweiligen Skala umgekehrt.

Um die erste Forschungsfrage (F1) zu beantworten, werden einerseits bayesianische *t*-Tests mit den sieben Globalurteilen zur Erklärqualität der beiden Statusgruppen Lehrkräfte und Didaktiker*innen gerechnet. Die Analyse erfolgt mit dem R-Package *Bayes-Factor* (Morey et al., 2022).¹⁰⁰ Anhand des zu bestimmenden Faktors (Bayes Faktor, BF10) wird im weiteren Verlauf angegeben, wie wahrscheinlich die H0 (beide Statusgruppen unterscheiden sich nicht signifikant) gegenüber der H1 (beide Statusgruppen unterscheiden sich signifikant) angenommen werden kann. Nach Schönbrodt und Wagenmakers (2017) entsprechen Werte unterhalb des Wertes 1 – dieser besagt, dass beide Hypothesen gleich wahrscheinlich sind – für die H0 und oberhalb für die H1.¹⁰¹ Für einen Vergleich der kriterienbasierten Beurteilungen beider Statusgruppen werden andererseits Profilvergleiche beider Gruppen erstellt (siehe z. B. Geiser & Eid, 2006, S. 326–328) und Abweichungen der Summenscores pro Merkmal anhand eines *Chi-Quadrat*-Tests auf Signifikanz untersucht. Erwartet wird auch hier, dass keine signifikanten Unterschiede vorliegen, weshalb von einem niedrigen X^2 -Wert ausgegangen wird.

Für die Forschungsfragen zwei (F2), drei (F3) und vier (F4) werden zweifaktorielle Varianzanalysen mit Messwiederholung durchgeführt (Leonhart, 2015, Kap. 19; *mixed ANOVA*).¹⁰² Die *mixed ANOVA* wird mittels des R-Pakets *ez* (Lawrence, 2016) berechnet, die erforderliche Umstrukturierung des Datensatzes in ein Langformat erfolgt mittels des R-Pakets *multilevel* (Bliese, 2022). Als abhängige Variable dient das Globalurteil zur Erklärqualität; Statusgruppe und Erklärmodus gehen als unabhängige Variablen in die Analyse ein. Abhängige Daten liegen vor, da dieselben Personen verschiedene Erklärungen zu einem Thema bewerten, wobei Erklärmodus im linearen Modell als Messwiederholungsfaktor fungiert. Die erforderliche Überprüfung der Voraussetzungen erfolgt für die Annahme der Normalverteilung mittels *Kolmogorov-Smirnov*-Test. Bühner und Ziegler (2017, S. 561–562) empfehlen bei annähernd gleich großen Stichproben auch bei Verletzung der Normalverteilungsannahme die Berechnung von Varianzanalysen, da dennoch

99 Ein Beispiel ist das Item SV_4 (siehe Tabelle 4.6): „Manche Sätze hat der Lehrer zu lang gemacht“.

100 Wenn nicht anders beschrieben erfolgen alle Analysen mit Hilfe des R-Pakets *base* (R Core Team, 2021).

101 Für genaue Abstufungen innerhalb der Bereiche ober- und unterhalb von 1, siehe Table 1 bei Schönbrodt und Wagenmakers, 2017, S. 130.

102 Möglich wäre zur Beantwortung dieser Frage auch, das *Linear Mixed Regression Model* nach Hilbert et al. (2019) einzusetzen, da dieses unter anderem Vorteile hinsichtlich statistischer Power aufweist. Um Vergleichbarkeit mit den anderen Analysen zu FALKE-q zu gewährleisten (siehe Frei et al., i. Vorb.; siehe Schilcher et al., i. Vorb.) wird in dieser Arbeit dennoch das Verfahren der *mixed ANOVA* durchgeführt.

robuste Ergebnisse zu erwarten sind. In *Anhang D1* werden die Ergebnisse aller *Kolmogorov-Smirnov*-Tests zusammengefasst. Zur Überprüfung der Sphärizitätsannahme wird der *Mauchly*-Test eingesetzt (siehe *Anhang D3*); bei Verletzung dieser Voraussetzung wird eine *Greenhouse-Geisser*-Korrektur der Freiheitsgrade vorgenommen. Die Voraussetzung homogener Varianzen in den untersuchten Teilgruppen wird mithilfe des *Levene*-Tests überprüft (R-Paket *car*, Fox et al., 2022). Bei Verletzung dieser Annahmen können bei annähernd balancierten Designs und größeren Stichproben dennoch robuste Ergebnisse erwartet werden (Bortz & Schuster, 2010, S. 214). Eine potenzielle Überschätzung der Effekte wird an dieser Stelle aus Gründen der Vergleichbarkeit aller Ergebnisse in Kauf genommen, limitiert jedoch die Aussagekraft. Eine Übersicht aller Tests auf Varianzhomogenität wird ebenfalls in *Anhang D2* gegeben.

Für die Beantwortung von Forschungsfrage zwei (F2) werden die beiden Faktoren, Statusgruppe und Erklärmodus, sowie alle potenziellen Interaktionen im Modell berücksichtigt. Mittels *Post-hoc*-Analysen fokussiert die dritte Forschungsfrage 3 (F3) den Faktor Statusgruppe¹⁰³ und die vierte Forschungsfrage (F4) den Faktor Erklärmodus. Die Signifikanzniveaus der dabei eingesetzten klassischen *t*-Tests sind *Bonferroni*-korrigiert, um der Kumulierung der Alphafehler zu begegnen.

Anhand der fünften Forschungsfrage (F5) soll beantwortet werden, wie die Bewertung der Erklärqualität mit den Beurteilungen zu einzelnen Qualitätsmerkmalen unterrichtlichen Erklärens zusammenhängt. Angelegt wird dafür ein Mehrebenenmodell (hierarchisch lineares Regressionsmodell) mit der abhängigen Variable Globalbewertung (Erklärqualität) sowie sechs Prädiktoren, bestehend aus den fünf Qualitätsmerkmalen und der Kontrollvariable Persönlichkeitswirkung. Die Daten sind dabei sowohl nach Personen (intraindividueller Faktor) als auch nach Videos (interindividueller Faktor) strukturiert. Die Analyse erfolgt mit dem R-Paket *lme4* (Bates et al., 2022). Als Maß für den Einfluss dieser *random effects* wird die Intraklassenkorrelation (kurz *ICC*) angegeben (Field, 2013, S. 729). Pro Statusgruppe wird ein eigenes Modell berechnet, sodass insgesamt drei Modelle vorliegen.¹⁰⁴

Zur Beantwortung der siebten Forschungsfrage (F7)¹⁰⁵, inwieweit das Vorwissen der Schüler*innen mit deren Bewertung der Erklärqualität zusammenhängt, kommt das Linear Mixed Regression Model nach Hilbert et al. (2019) zum Einsatz. Die abhängige Variable ist das Globalurteil, während Personen und Videos als *random effects* fungieren und der Gesamtscore im Vorwissenstest sowie der Erklärmodus mittels Dummy-Codierung als *fixed effects* in das Modell eingehen. Für die Analyse wird ebenfalls das R-Paket *lme4* (Bates et al., 2022) verwendet.

103 Um F4 umfänglich beantworten zu können, müssen außerdem die offenen Antworten inhaltsanalytisch ausgewertet werden (F4b). Die dabei zum Einsatz kommende qualitative Inhaltsanalyse wird in Kapitel 4.9.2 erläutert.

104 In einem anderen Forschungsbericht werden zwei Modelle gerechnet: Ein Modell für die Schüler*innen und ein Modell, das Studierende und Lehrende zusammenfasst (siehe Frei et al., i. Vorb.). Die Darstellung erfolgt dort aus Gründen der fachübergreifenden Vergleichbarkeit.

105 F6 bezieht sich ausschließlich auf die Textdaten. Die inhaltsanalytische Auswertung dazu wird in Kapitel 4.9.2 beschrieben.

4.9.2 Qualitative Inhaltsanalyse

Die qualitativen Daten der vorliegenden Studie wurden im Rahmen des Onlinefragebogens von Studierenden und Lehrenden in Form offener Antwortmöglichkeiten erhoben, wobei die Beantwortung nicht verpflichtend war. Da nicht alle der 123 Personen von dieser Option Gebrauch machten, liegen über alle Teilnehmenden keine vergleichbaren qualitativen Daten vor. Aussagen in Form offener Antworten waren an drei Stellen des Fragebogens möglich, die Erhebung erfolgte computergestützt. Etwaige Tipp- oder Rechtschreibfehler sowie Mängel in Satzbau oder Ausdruck wurden nicht korrigiert, sodass die originalen Antworten vorliegen. Für den Ergebnisbericht wurden die Daten anonymisiert, indem die Teilnehmer-Codes (tn_code) entfernt und die einzelnen Antworten pro Video für Studierende sowie für verschiedene Untergruppen der Lehrenden (*Prof*: Professor*innen; *Hsl*: Hochschullehrende; *Slk*: Seminarlehrkräfte; *Lk*: Lehrkräfte; *Stud*: Studierende) nummeriert wurden. Es werden keine Fallvergleiche zwischen einzelnen Personen vorgenommen, die Nummerierung erfolgt für jede Statusuntergruppe gesondert. Lehrkraft 7 von Video A (z. B. *Lk.7_NW.o_Bg*)¹⁰⁶ entspricht demnach Lehrkraft 7 bei Video B (z. B. *Lk.7_TL.aa_Bg*). Seminarlehrkraft 7 ist bei Video A jedoch eine andere (z. B. *Slk.7_NW.o_Bg*). Der gesamte Datenkorpus ist unter *Anhang G* zu finden und umfasst 135.933 Zeichen. Für die Analyse wurde die Software *MAXQDA*¹⁰⁷ verwendet.

Zur Einteilung des Datenmaterials folgt die vorliegende Arbeit der Terminologie von Kuckartz (2016, Kap. 2): Als Auswahlinheit dienen jeweils alle Antworten einer Statusuntergruppe (Professor*innen, Hochschullehrende, Seminarlehrkräfte, Lehrkräfte und Studierende)¹⁰⁸ zu einer Erklärung. Demnach liegen bei sieben Erklärungen und den fünf Gruppen insgesamt 35 Auswahlheiten vor. So können nicht nur Antworteigenschaften der unterschiedlichen Statusgruppen berücksichtigt werden. Auch dient die Einteilung dazu, beim Codieren einzelne Kategorien zu fokussieren, die aufgrund der systematischen Variation von Erklärthema und Erklärmodus entweder nur für ein bestimmtes Thema oder im Kontext einer konkreten Repräsentationsform vorliegen. Eine weitere Unterscheidung in Analyseeinheiten ist nicht notwendig (siehe Kuckartz, 2016, S. 30–31). Als Codiereinheit¹⁰⁹ wird eine „Textstelle verstanden, die mit einer bestimmten Kategorie [...] in Verbindung steht“ (Kuckartz, 2016, S. 41). Es wurden keine Textsegmente für das Codieren vorgegeben, der Umfang der Codierungen reicht von einzelnen Wörtern bis hin

106 Die Abkürzung bedeutet: Lehrkraft Nummer 7, Erklärung NW_o, Textfeld „Begründung“. Andere Abkürzungen zu den Statusuntergruppen (Prof, Hsl, Slk und Stud) sind erklärt. Zudem steht „Vb“ für das Textfeld Verbesserungen und „Rf“ für Aussagen zu den einzelnen Repräsentationsformen. Weitere Beispiele könnten also sein: Prof.4_TA.a_Vb; Stud.18_TA.aa_Rf.

107 Siehe <https://www.maxqda.de>.

108 Diese Unterscheidung wurde getroffen, um mögliche Unterschiede (z. B. gruppenspezifisches Antwortverhalten wie Ausführlichkeit der Begründungen oder das Unterbreiten von Vorschlägen) zwischen diesen Gruppen bereits im Analyseprozess festzustellen. Bei der Analyse zeigte sich, dass in der Ergebnisdarstellung eine Unterscheidung zwischen Studierenden und Lehrenden (hier Professor*innen, Hochschullehrende, Seminarlehrkräfte und Lehrkräfte) sinnvoll ist.

109 In Anlehnung an Kuckartz (2016) werden synonym auch die Begriffe *codiertes Segment* oder *Textsegment* verwendet.

zu mehreren Sätzen. Damit fallen in der vorliegenden Studie die Prozesse des Segmentierens und des Codierens zusammen.

Mit der Erhebung und Auswertung der qualitativen Daten verfolgt die vorliegende Studie mehrere Ziele und möchte damit insbesondere die sechste Forschungsfrage (F6) beantworten. Erstens soll anhand der Begründungen zur Bewertung der Erklärqualität festgestellt werden, ob die theoretisch hergeleiteten und ab dem zweiten Durchgang des Fragebogens quantitativ erhobenen Merkmale guten Erklärens durch diese Daten validiert werden können (F6a). Da beim ersten Durchgang neben den Bewertungen der Qualität aller sieben Erklärungen noch keine Beurteilungen hinsichtlich der ausgewählten Merkmale erfolgt, kann deren Relevanz anhand der Nennung in den offenen Antworten im Sinne einer Konstruktvalidierung überprüft werden.

Zweitens sollen die Textantworten der Studienteilnehmer*innen auch dazu dienen, neben den theoretisch postulierten weitere potenzielle Qualitätsmerkmale von Erklären im Musikunterricht zu erfassen und damit blinde Flecken auszuleuchten, die von den geschlossenen Items übersehen werden (F6b). Ein besonderer Fokus der vorliegenden Arbeit richtet sich auf unterschiedliche Erklärmodi, die sich aus dem Einsatz verschiedener Formen der Repräsentation ergeben können. Um Forschungsfrage vier umfassender zu beantworten, können die qualitativen Daten insofern Erkenntnisse liefern, als auch Äußerungen über den generellen Einsatz optischer oder akustischer Repräsentationen in den offenen Antworten identifiziert und die statistischen Befunde (F4a) durch einen anderen forschungsmethodischen Zugang ergänzt werden können (F4b).

Inhaltlich strukturierende qualitative Inhaltsanalyse

Die vorliegende Untersuchung folgt grundsätzlich dem Schema der inhaltlich strukturierenden qualitativen Inhaltsanalyse nach Kuckartz (2016, Kap. 5). Dieses Verfahren erschien nicht nur deshalb geeignet, weil eine inhaltliche und themenorientierte Auswertung von Textmaterial im Vordergrund steht, sondern auch, da es das Datenmaterial anhand von zwei Dimensionen strukturieren kann: Fälle und Kategorien. Der Fokus liegt vor allem auf den Kategorien, da insbesondere eine themenorientierte Auswertung des Materials von Interesse ist. Fallbezogene Auswertungen scheinen aufgrund fehlender Vollständigkeit und Balance zwischen den Gruppen sowie wegen der teils sehr knappen Antworten nicht sinnvoll möglich.

Das Vorgehen bei der Kategorienbildung erfolgt deduktiv-induktiv. Zum einen liegen aufgrund der theoretisch postulierten Qualitätsmerkmale bereits deduktive Kategorien vor, die *a priori* festgelegt wurden und eine enge Verbindung zur theoretisch-inhaltlichen Rahmung der Untersuchung herstellen. Zum anderen erfordert insbesondere der zweite Teil von Forschungsfrage sechs (F6b), welche weiteren Qualitätsmerkmale für unterrichtliches Erklären im Fach Musik in den Begründungen identifiziert werden können, eine induktive Kategorienbildung am Material. An dieser waren beide Personen beteiligt, die den Textkorpus codiert haben: der Autor der vorliegenden Arbeit und eine beteiligte Hilfskraft, die Lehramt Musik an Gymnasien sowie Gesangspädagogik studiert hat. Im Zentrum der Kategorienbildung stand die Strukturierung und Aufschlüsselung des

Datenmaterials hinsichtlich der zu beantwortenden Forschungsfragen (siehe Kuckartz, 2016); die Entwicklung eines übertragbaren Analyseinstrumentariums war von sekundärem Interesse.

Der Analyseprozess kann in zwei Abschnitte unterteilt werden, denen ein unterschiedliches Kategoriensystem zugrunde lag. Um die Transparenz im methodischen Vorgehen zu gewährleisten, wird der gesamte Prozess linear beschrieben und beginnt demnach mit dem ersten Kategoriensystem (*Kategoriensystem 1a*, Anhang F1), das jedoch nach der Codierung des gesamten Materials (siehe *Kategoriensystem 1b*, Anhang F2) und ersten Schritten der Auswertung wieder verworfen wurde. Mit einem zweiten Kategoriensystem (*Kategoriensystem 2a*, siehe Anhang F3) erfolgte ein weiterer Komplettdurchlauf durch das Material. Da beide Codierer*innen im Verlauf des konsensuellen Codierens weitere Anpassungen der Kategorien vorgenommen hatten, wird das endgültige und der späteren Analyse (siehe Kap. 5) zugrunde liegenden Manual als *Kategoriensystem 2b* (siehe Anhang F4) bezeichnet.

Kuckartz (2016) beschreibt die inhaltlich strukturierende qualitative Inhaltsanalyse als „mehrstufiges Verfahren der Kategorienbildung und Codierung“ (S. 97) und weist dafür sieben Prozessschritte aus (siehe Kuckartz, 2016, S. 100, Abb. 16). Das methodische Vorgehen bei der Analyse der qualitativen Daten orientiert sich an dieser Struktur, entsprechende Abweichungen werden benannt.

Verfahrensdokumentation

Da sich gerade die Datenerhebung bei der Gruppe der Lehrenden über den Zeitraum von mehr als einem Jahr erstreckte, erfolgte die *initiiierende Textarbeit* in der ersten Phase parallel zur Datenerhebung (Kuckartz, 2016, S. 56–57). Während auf Fallzusammenfassungen aufgrund geringer Daten pro Studienteilnehmer*in beziehungsweise unvollständiger Daten pro Statusgruppe verzichtet wurde, dienten kurze Notizen zu inhaltlichen Auffälligkeiten als Memo (siehe Kuckartz, 2016, S. 57–58). Beispielsweise wurde notiert, dass insbesondere bei den Erklärungen TA_a (Video C) und TL_a (Video G) das Fehlen der optischen Repräsentation auffallend häufig kritisiert und dementsprechend auch negativ bewertet wurde. In dieser Phase wurde für die Memos nicht, wie von Kuckartz (2016) vorgeschlagen, mit *QDA-Software* (Qualitative Data Analysis Software) gearbeitet, sondern mittels Word-Dokument.

Kategoriensystem 1a

In der zweiten Phase stand die „Entwicklung thematischer Hauptkategorien“ (Kuckartz, 2016, S. 101) im Vordergrund. Für das *Kategoriensystem 1a* (siehe Anhang F1) lagen acht deduktive Hauptkategorien vor. Diese umfassten die theoretisch postulierten Qualitätsmerkmale unterrichtlichen Erklärens sowie die Kontrollvariable zur Persönlichkeit der Lehrkraft: (1) Adressatenorientierung, (2) Strukturiertheit, (3) Sprachliche Verständlichkeit, (4) Sprech- und Körperausdruck sowie jeweils die (5) optische, die (6) akustische

und die (7) verbale Repräsentationsform und die (8) Sympathie der Lehrkraft. Mit diesen deduktiven Hauptkategorien startete die erste Codierung durch den Autor an etwa 25 % der Videos NW_o (Video A), TA_a (Video C) und TL_oa (Video B). So war gewährleistet, dass durch die drei Videos sowohl jedes Erklärthema als auch jeder Erklärmodus abgedeckt war.

Während dieses ersten Codierdurchgangs kamen induktive Kategorien wie der Einsatz von Analogien oder die Forderung nach praktischer Umsetzung mit den Schüler*innen dazu (siehe die hellgrau hinterlegten Kategorien in *Kategoriensystem 1a, Anhang F1*). Außerdem wurden In-Vivo-Codierungen nach einzelnen Wörtern, wie *abstrakt*, *verständlich* oder *nachvollziehbar* vorgenommen, die das Kategoriensystem ergänzt haben. Gelöscht wurde die Kategorie *verbale Repräsentationsform* (siehe die dunkelgrau hinterlegte Kategorie in *Kategoriensystem 1a, Anhang F1*), da diese im Vergleich zu den Kategorien *Sprachliche Verständlichkeit* und *Adressatenorientierung* nicht trennscharf war. Umbenannt und inhaltlich geweitet wurde die Kategorie zur Kontrollvariable *Sympathie*; diese wurde in *Lehrkraft* umbenannt und bezieht sich neben der Persönlichkeit auch auf andere Aspekte wie das Aussehen der Lehrkraft.

Gerade anhand der induktiv gewonnenen Kategorien zeigte sich, dass nicht nur eine weitere Ausdifferenzierung des Manuals, sondern auch eine Hierarchisierung der Kategorien notwendig waren. Beispielsweise wurde deutlich, dass der Code *Musikbezug* der Kategorie *Inhaltliches zur Erklärung* untergeordnet werden muss. Allerdings gab es problematische Textstellen, die nach Diskussion zwischen beiden codierenden Personen nicht zufriedenstellend codiert werden konnten. Die Mängel am Kategoriensystem führten dazu, dass weder die Übereinstimmung beider Codierer*innen überprüft noch das gesamte Material mit den Hauptkategorien codiert wurde (siehe Kuckartz, 2016). Im weiteren Verlauf wurde ein Kategoriensystem mit neuer Struktur in Gestalt von Haupt- und Subkategorien entwickelt (siehe *Kategoriensystem 1b, Anhang F2*), in dem die bestehenden Codes aus *Kategoriensystem 1a* integriert und gewonnene Erfahrungen aus dem ersten Codierdurchgang berücksichtigt wurden. Mit Blick auf das Phasenmodell nach Kuckartz (2016) erfolgte in diesem Schritt nicht nur das Anlegen thematischer Haupt-, sondern bereits auch Subkategorien.¹¹⁰

Kategoriensystem 1b

Die Struktur von *Kategoriensystem 1b* beinhaltete sechs Hauptkategorien: (1) Erklärung, (2) Schüler*innen, (3) Lehrkraft, (4) Thema, (5) Studiendesign und (6) Sonstiges. Diese unterscheiden sich in ihrem Umfang teils beträchtlich, da beispielsweise unter der Hauptkategorie *Erklärung* insgesamt 88 weitere Kategorien auf mehreren Ebenen zusammengefasst sind, während sich die Kategorien *Lehrkraft* aus zwei oder *Thema* aus drei Subkategorien einer Ebene zusammensetzen. Die nächsttiefere Ebene der Hauptkategorie

¹¹⁰ Da die dritte Phase des Modells aufgrund der begrenzten zeitlichen Ressourcen bereits adaptiert wurde, konnte auch Phase vier, das „Zusammenstellen aller mit der gleichen Hauptkategorie Codierten Textstellen“ (Kuckartz, 2016, S. 100) nicht durchgeführt werden, sondern wurde ebenfalls um Subkategorien erweitert.

Erklärung besteht aus zehn weiteren Kategorien (siehe *Anhang F2*). Deutlich wird dabei der Schwerpunkt auf den zentralen Untersuchungsgegenstand der Studie. Abweichend von der Empfehlung Kuckartz' (2016, S. 102–105) wurde aus zeitökonomischen Gründen kein kompletter Durchlauf am Material lediglich anhand der Hauptkategorien vorgenommen, sondern es wurden direkt auch Subkategorien mitcodiert. Dafür lagen vor dem Durchgang insgesamt 76 Kategorien vor. Während des Codierens durch das gesamte Material kamen noch weitere 44 induktive Kategorien dazu und vier wurden am Ende des Gesamtdurchgangs gelöscht. Das Codieren des Materials erfolgte nach Video beziehungsweise Erklärthema, sodass mehrere Absprachen während des Codierdurchgangs möglich waren. Als Textkorpus wurden alle Antworten definiert, die als Begründung zur Bewertung der Erklärqualität im ersten Videodurchgang abgegeben wurden. Dabei entstanden 1.695 Codierungen von Person 1 und 1.658 Codierungen von Person 2.

An dieser Stelle hätte mit den nun vorliegenden 120 Kategorien Phase 6 begonnen, die einen Komplettdurchlauf beider Codierer*innen durch das gesamte Material vorsieht (Kuckartz, 2016, S. 110–111). Jedoch erwiesen sich insbesondere zwei Aspekte als problematisch: Erstens orientierten sich die Kategorien zum Einsatz von Repräsentationen insbesondere daran, ob sie in der Erklärung vorhanden waren oder fehlten, und daran, inwieweit dies als positiv oder negativ bewertet wurde. Gerade die Bewertung erforderte zwischen den Textstellen ein unterschiedliches Maß an Interpretation. Folgendes Beispiel soll dies verdeutlichen: Video TL_a wurde von einer Lehrkraft mit der Note 3 bewertet. Als Begründung wurde die „fehlende Visualisierung“ (*Lk.6_TL.a_Bg*) genannt. Es ist hier nicht zu entscheiden, ob die Note aus der fehlenden Visualisierung resultiert oder ob es mit einer Visualisierung eine schlechtere Note gegeben hätte. Die erstgenannte Variante erscheint plausibel, erfordert aber Interpretation. Zudem wurden bei den Repräsentationsformen unterschiedliche Aussagen in einer Kategorie vereint: Für den Code *ARF-vorhanden* wurden sowohl Textstellen wie „besser, da anhand des Höreindrucks erklärt wird“ (*Slk.6_TL.a_Bg*) als auch „man hört die Betonung im Original nur sehr schlecht“ (*Lk.22_TA.aa_Bg*) codiert. Während das erste Textbeispiel eine allgemeine Aussage zum Einsatz akustischer Repräsentation darstellt, bezieht sich das zweite auf die spezifische Umsetzung der akustischen Repräsentation in der Erklärung.

Der zweite problematische Aspekt zeigte sich beim Blick auf die Struktur des *Kategoriensystems Ib*: Die Textsegmente sollen insbesondere auf Aussagen zu Qualitätsmerkmalen unterrichtlichen Erklärens untersucht werden: Diese verteilen sich jedoch über verschiedene Hauptkategorien (z. B. Strukturiertheit bei Hauptkategorie *Erklärung*, Adresatenorientierung bei Hauptkategorie *Schülerinnen und Schüler* oder Sprech- und Körperausdruck bei Hauptkategorie *Lehrkraft*). Zudem werden die Aussagen zu Repräsentationsformen im Kategoriensystem umfassender aufgegliedert, orientieren sich jedoch nach dem Fehlen beziehungsweise Vorkommen dieser. Inhaltliche Aussagen hingegen, wie „Der Zeitaspekt würde deutlicher werden, wenn die Notenwerte entsprechend ihrer Länge direkt untereinander stünden und nicht wie eine Pyramide aufgebaut wären“ (*Slk.8_NW.o_Bg*), sind unter dem Code *Empfehlungen* zusammengefasst. Das Kategoriensystem strukturiert sich demnach eher nach der Bedeutung der Aussagen und weniger nach den Qualitätsmerkmalen. Wird also eher angesprochen, ob eine optische

Repräsentation vorkommt, oder werden Empfehlungen zum Einsatz optischer Repräsentationen hinsichtlich guten Erklärens gegeben? Vor diesem Hintergrund wurde das *Kategoriensystem 1b* nochmals modifiziert. Aufgrund der veränderten Struktur wird von *Kategoriensystem 2a* gesprochen.

Kategoriensystem 2a

Das neue *Kategoriensystem 2a* ist in drei Bereiche unterteilt: (1) Überzeugungen zum Einsatz von Repräsentationen, (2) Konzept der Erklärung und (3) Performanz beim Erklären. Darüber hinaus liegen vier Kategorien vor, die nicht speziell auf einzelne Fragestellungen abgestimmt sind, sondern weitere Aspekte betreffen, die häufig im Datenmaterial zu finden waren. In der Kategorie *Keine Qualitätsmerkmale* sind Aussagen zusammengefasst, die nichts mit Erklären oder potenziellen Qualitätsmerkmalen zu tun haben (z. B. „Medientechnisch aufwändig, aber umständlich“, *Slk.1_NW.oa_Bg*). Häufig werden die Worte *Theorie* und *Praxis* als Begründung für die Bewertung der Erklärqualität aufgeführt (z. B. „Kombination von Theorie und Praxis“, *Stud.25_TL.oa_Bg*; „Praxis fehlt“, *Stud.15_NW.o_Bg*). Allerdings erfolgt die Codierung in einer gesonderten Kategorie, da weder ein eindeutig identifizierbares noch einheitlich verwendetes konzeptuelles Begriffsverständnis herauszuarbeiten war. Zudem markierte die Kategorie *Interessant* Aussagen, die aus Sicht der Codiererin sowie des Codierers hinsichtlich der Forschungsfragen besonders interessant erschienen. Die Kategorie *SuS-Aktivität* (z. B. „allerdings fehlt weiterhin die Schüleraktivität (Klatschen etc.)“, *Slk.6_NW.oa_Bg*) bezieht sich auf Aussagen, dass die Schüler*innen mehr eingebunden werden müssen und selbst aktiv klatschen oder singen sollen – dies war jedoch aufgrund des Studiendesigns nicht möglich.

Bereich eins umfasst vier Kategorien, in denen grundsätzliche Überzeugungen der Studienteilnehmer*innen zum Einsatz von Repräsentationen zusammengefasst sind. Zur Beantwortung des dritten Teils von Forschungsfrage vier (F4c) wurden Textstellen codiert, in denen neben der Verwendung von Analogien insbesondere der Einsatz optischer, akustischer beziehungsweise beider Repräsentationsformen in Kombination als Begründung für die Bewertung der Erklärqualität genannt wurde. Allen Aussagen wie „Visualisierung fehlt“ (*Hsl.1_TA.a_Bg*) oder „Höreindruck fehlt“ (*Slk.7_TL.o_Bg*) ist gemein, dass weder eine Wertung im Sinne von „Tafelbild würde helfen“ (*Stud.62_Ta.a_Vb*) vorgenommen wird. Noch beziehen sich die Aussagen auf die konkrete Umsetzung in der Erklärung. Codiert werden Äußerungen zum Fehlen oder Vorkommen einer optischen oder akustischen Repräsentation, die jedoch nicht näher spezifiziert sind.

Deutlich umfangreicher in der Anzahl der Kategorien ist der Bereich *Konzept*. Dieser umfasst neun Hauptkategorien und 45 Subkategorien, mit denen codiert wurde. Neben themenspezifischen Kategorien wie *Klavatur* (Erklärthema: Tonleiter) oder *Rhythmus-Verfremdung* (Erklärthema Taktarten) sowie Aussagen zum *Inhalt* (Kategorie 3.5) sind insbesondere Äußerungen zu den ausgewählten Qualitätsmerkmalen unterrichtlichen Erklärens (Kategorie 3.6 bis 3.9) für die vorliegende Untersuchung und die Beantwortung des ersten Teils von Forschungsfrage sechs (F6a) relevant. Letzte unterscheiden sich im

Gegensatz zu den vorher genannten Überzeugungen, indem Aussagen codiert werden, die eine Wertung vornehmen und sich direkt auf die konkrete Umsetzung in der Erklärung beziehen (z. B. „Unterstützung durch Klangbeispiele ist gut“, *Stud.60_NW.oa_Bg*; „Auf den hinkenden Vergleich mit der Stunde wird verzichtet“, *Slk.4_NW.oa_Bg*). Weitere Kategorien sind in *Anhang F3* aufgeführt.

Der dritte Bereich *Performanz* umfasst Aussagen zur Lehrkraft sowie ihrem Handeln in der Erklärung und subsummiert fünf Kategorien. Zentral sind dabei Textstellen zum Qualitätsmerkmal *Sprech- und Körperausdruck*, die sich auf die drei Kategorien *Sprechen*, *Gestik* und *Zugewandtheit/Blickkontakt* aufteilen. Weitere Kategorien sind *Persönlichkeit* (z. B. „am Anfang Humor [*der Lehrkraft*, vom Autor hinzugefügt] ist gut“, *Stud.47_NW.o_Bg*) sowie *Professionskönnen* – gemeint sind damit beispielsweise Aussagen über Fertigkeiten beim Vorspielen am Klavier (z. B. „Vorspielen ungenau“, *Stud.32_NW.oq_Bg*).

Mit diesem *Kategoriensystem 2a* erfolgte der Gesamtdurchgang durch das Material, innerhalb dessen 16 weitere Kategorien induktiv gewonnen worden sind (grau hinterlegt, siehe *Anhang F3*). Der Gesamtdurchgang erfolgte video- und themenweise in dieser Reihenfolge: *NW_oa*, *NW_o*, *TA_oa*, *TA_a*, *TL_oa*, *TL_o* und *TL_a*. Aus zeitlichen wie forschungsökonomischen Gründen verliefen die Phasen fünf und sechs des Modells von Kuckartz (2016, S. 100) parallel. Wenn beispielsweise im Textsegment von Erklärung *TA_a* eine neue Kategorie hinzugekommen ist, wurde mit dieser nochmal das Material der vorher codierten Videos (*NW_oa*, *NW_o* und *TA_oa*) durchgearbeitet. Dabei ergaben sich in den allermeisten Fällen jedoch keine Änderungen, da die Kategorien meist themenbezogen waren und daher nur die Textstellen desselben Themas kleine Änderungen hervorriefen. Mit Ausnahme der Kategorien *B_allgemeine Bewertung* und *B_(Miss-)Verständlichkeit* weisen alle neu hinzugekommenen Kategorien nur wenige codierte Stellen auf. Die beiden genannten Kategorien beziehen sich auf Aussagen zu verwendeten Beispielen (*B_*), die insbesondere bei den Erklärungen zu Taktarten in den offenen Antworten zu finden waren und sich dort häufig auf die Passung des verwendeten Rhythmusbeispiels beziehen (z. B. „einzig, der verwendete Rhythmus ist unpassend, da mit Überbindungen“, *Lk.3_TA.oa_Bg*).

Eine neue Kategorie wurde notwendig, da es weder um die optische Repräsentation geht – die kritisierte Überbindung ergab sich aus der Beschaffenheit des Rhythmus’ sowie der gewählten Taktart – noch um die akustische Darstellung. Am Ende des Gesamtdurchgangs wurden zusätzlich zu den neuen 16 Kategorien noch gezielte Begriffe wie „abstrakt“ (*Prof.6_TA.a_Bg*), „verständlich“¹¹¹ (*Lk.15_TL.o_Bg*) oder „anschaulich“ (*Prof.7_NW.o_Bg*) codiert, sodass weitere sechs Kategorien dazukamen.

111 Bei den Begriffen sind immer auch negative Verwendungen wie „nicht verständlich“ (*Slk.9_NW.oa_Bg*) zu codieren.

Kategoriensystem 2b

Das endgültige Kategoriensystem (siehe *Kategoriensystem 2b*, siehe *Anhang F4*) kann demnach in vier große Bereiche unterteilt werden: (1) Überzeugungen, (2) Konzept der Erklärung, (3) Performanz und (4) Begriffe. Die Kategorie *SuS-Aktivität* wurde im Verlauf inhaltlich Bereich eins zugeordnet, sodass *Überzeugungen* durch fünf Kategorien repräsentiert ist. Der zweite Bereich *Konzept* unterteilt sich weiterhin in neun Hauptkategorien, die in der endgültigen Version insgesamt 56 Subkategorien subsumieren. Der Bereich *Performanz* wurde um die Kategorie *Raumregie* (z. B. „keine ideale Raumposition der Lehrkraft“, *Stud.38_NW.o_Vb*) ergänzt und umfasst insgesamt sechs Kategorien. Bereich vier *Begriffe*, der neu dazugekommen ist, weist acht Kategorien auf, von denen eine bereits im *Kategoriensystem 2a* bestand: *Theorie/Praxis* wurde übernommen. Die beiden restlichen Kategorien *Interessant* und *Keine Qualitätsmerkmale* blieben ebenfalls bestehen.

Abschließend erfolgt ein Überblick zu Anzahl und Umfang der offenen Antworten. Die Darstellung trennt dabei nach Statusgruppe und den drei Positionen im Fragebogen (Begründung, Durchgang 1; Verbesserung, Durchgang 3; Repräsentationsformen, Durchgang 3), da dies für die Beantwortung des ersten Teils von Forschungsfrage sechs (F6a) relevant ist. Tabelle 4.12 zeigt, dass von insgesamt 1.004 offenen Antworten – unabhängig des Umfangs der Antwort wird jeweils nur gezählt, ob das Antwortfeld ausgefüllt wurde oder nicht – 575 (57,27 %) auf die Begründung zur Bewertung der Erklärqualität entfallen, die im ersten Durchgang erhoben wurde.

Tabelle 4.12 Anzahl der offenen Antworten über alle Videos

Offene Antworten	<i>Begründung</i>	<i>Verbesserung</i>	<i>Rep.formen</i>	Gesamt
Lehrende	301 29,98 %	115 11,45 %	108 10,76 %	524 52,19 %
Studierende	274 27,29 %	100 9,96 %	106 10,56 %	480 47,81 %
Gesamt	575 57,27 %	215 21,41 %	214 21,32 %	1.004 100 %

Während des dritten Durchgangs gab es zwei Optionen für eine Stellungnahme in Form einer offenen Antwort: Zum einen wurden 215 (21,41 %) Vorschläge zur generellen Verbesserung der Erklärung abgegeben (vgl. Kap. 4.6.2, Tab 4.10, Kv_3_aku, Kv_3_opt). Zum anderen tätigten die Studienteilnehmer*innen 214-mal (21,32 %) Aussagen zur Qualität der Repräsentationsformen, die bei der Erklärung zum Einsatz gekommen waren (siehe Kap. 4.5.2, Tab 4.5, MR_ver_1n, MR_opt_1n, MR_aku_1n). Die Möglichkeit der offenen Antworten wurde von Studierenden und Lehrenden etwa in gleichem Maße in Anspruch genommen.

Wie bereits erwähnt, unterscheiden sich die offenen Antworten teilweise extrem hinsichtlich ihrer Länge: Der Umfang reicht von einem Wort (Minimumwert: 7 Zeichen, *Lk.26_NW.o_Bg*) bis hin zu mehreren Sätzen (Maximumwert: 944 Zeichen, *Lk.3_TL.a_Bg*). Analog dazu stellt Tabelle 4.13 nun eine Übersicht der Zeichenzahlen da.¹¹² Der Textkorpus umfasst demnach insgesamt 135.933 Zeichen. Dabei ist der prozentuale Anteil der Lehrenden im Vergleich zur Anzahl der offenen Antworten etwas größer. Das bedeutet, dass die Aussagen der Lehrenden umfangreicher und tendenziell ausführlicher ausfallen als die Antworten der Studierenden. Dieser Anstieg ist insbesondere auf die Begründung zur Bewertung der Erklärqualität und auf die generellen Vorschläge zur Verbesserung der Erklärung zurückzuführen. Vermutet wird, dass diese Unterschiede mit der höheren Expertise der Lehrenden zusammenhängen könnten. Diese Frage wird in der vorliegenden Untersuchung jedoch nicht beantwortet.

Tabelle 4.13 Anzahl der Zeichen in den offenen Antworten über alle Videos

Offene Antworten	<i>Begründung</i>	<i>Verbesserung</i>	<i>Rep.formen</i>	Gesamt
Lehrende	48.589 35,75 %	16.862 12,40 %	11.801 8,68 %	77.252 56,83 %
Studierende	38.266 28,15 %	10.791 7,94 %	9.624 7,08 %	58.681 43,17 %
Gesamt	86.855 63,90 %	27.653 20,34 %	21.425 15,76 %	135.933 100 %

Die finale Codierung, auf deren Basis die Analyse erfolgt, beinhaltet 3.370 codierte Stellen. Während sich dabei pro Video im Schnitt 481 Codierungen ergeben, liegen die wenigsten zu Video F (TA_oa; 425 Codierungen), die meisten zu Video G (560, TL_a) vor. 1.950 Codierungen ergeben sich auf Basis der Aussagen aus der Gruppe der Lehrenden, wobei davon 233 auf Professor*innen, 637 auf Hochschullehrende, 418 auf Lehrkräfte und 662 auf Seminarlehrkräfte entfallen. Von den Studierenden wurden insgesamt 1.420 Aussagen codiert.

¹¹² Es wäre auch möglich gewesen, nur Worte zu zählen. Allerdings enthalten manche Antworten auch Zeichen wie + oder -, anhand derer eine Wertung analog zu *positiv* und *negativ* vorgenommen wurde. Beim bloßen Zählen von Wörtern, wären diese nicht berücksichtigt worden.

5 Ergebnisse

5.1 Vorbereitende Analyse zur Zusammensetzung der Statusgruppen

F1) Unterscheiden sich die Bewertungen der Erklärqualität und die Beurteilungen zu Merkmalen guten Erklärens zwischen Lehrkräften und Didaktiker*innen signifikant?

Die Annahme, dass Lehrkräfte und Didaktiker*innen in der vorliegenden Studie als eine Gruppe von Lehrenden aufzufassen sind, wird in einer vorbereitenden Analyse¹¹³ überprüft und stützt sich vor allem auf drei Gründe:

Erstens liegt sowohl innerhalb der Lehrkräfte als auch innerhalb der Didaktiker*innen eine heterogene Zusammensetzung der Stichproben vor, die eine sinnvolle Unterscheidung zwischen beiden Gruppen erschwert. Während in der Gruppe der Lehrkräfte neben erfahrenen Seminarlehrkräften und Lehrkräften mit langjähriger Berufserfahrung auch Berufsanfänger kurz nach Abschluss des Referendariats vertreten sind, umfasst die Gruppe der Didaktiker*innen neben Professor*innen und promovierten Hochschullehrenden auch wenige Promovierende, die noch kein zweites Staatsexamen absolviert haben. Zweitens geht damit die Problematik einher, dass aufgrund der Heterogenität innerhalb der beiden Untergruppen keine theoretisch sinnvolle Aufteilung hinsichtlich des Expertisegrades in diesem Personenkreis möglich ist. Drittens fällt die Stichprobe der Didaktiker*innen in dieser Untersuchung sehr klein aus ($N_{did} = 21$), sodass eine Betrachtung als Teilstichprobe mit erheblichen Limitationen bei den nachfolgenden quantitativen Analysen verbunden und die Aussagekraft für diese Gruppe somit stark eingeschränkt wäre. Aufgrund der genannten Aspekte erscheint die Unterscheidung zwischen Studierenden als Noviz*innen vor Abschluss der universitären Berufsausbildung und Lehrenden (Lehrkräfte und Didaktiker*innen) als Expert*innen nach Abschluss dieser angemessen und wird für die vorliegende Untersuchung angenommen.

Dies wird anhand von zwei Verfahren inferenzstatistisch geprüft. Zum einen werden die sieben Globalurteile beider Gruppen mittels bayesianischer t -Tests dahingehend untersucht, wie wahrscheinlich entweder die Nullhypothese (beide Gruppen unterscheiden sich nicht) oder die Alternativhypothese (beide Gruppen unterscheiden sich signifikant) zutrifft. Mit Profilanalysen werden die Beurteilungen zu den Qualitätsmerkmalen anhand der arithmetischen Mittelwerte beider Gruppen miteinander verglichen, indem ein *Chi-Quadrat*-Test diese auf signifikante Unterschiede überprüft. Tabelle 5.1 zeigt die Ergebnisse der bayesianischen t -Tests.

¹¹³ In der Regel beginnt der Ergebnisteil mit der Darstellung der Gütekriterien. In dieser Arbeit werden jedoch zunächst die Ergebnisse der vorbereitenden Analyse zur Zusammensetzung der Statusgruppen berichtet. Sie sind Grundlage für eine valide Berechnung der internen Konsistenzen, die Teil des Abschnitts 5.2 sind.

Tabelle 5.1 Bayesianische *t*-Tests pro Video zwischen den Statusgruppen Lehrkräfte und Didaktiker*innen

Erklärthema	Notenwerte		Taktarten		Tonleiter		
	optisch	opt./aku.	akustisch	opt./aku.	optisch	akustisch	opt./aku.
Videokürzel	NW_o	NW_oa	TA_a	TA_oa	TL_o	TL_a	TL_oa
Statusgruppe	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Lehrkräfte (<i>N</i> _{max} = 41)	2,28 (0,97)	2,45 (1,07)	3,13 (1,16)	2,04 (1,05)	2,36 (1,06)	3,70 (1,18)	2,02 (1,09)
Didaktiker*innen (<i>N</i> _{max} = 21)	2,48 (1,00)	2,26 (0,86)	2,92 (1,32)	2,46 (1,07)	2,81 (0,90)	3,61 (1,33)	2,16 (0,89)
Bayesianische <i>t</i>-Tests unabhängiger Stichproben							
BF ₁₀	0,343	0,336	0,327	0,656	0,827	0,284	0,305

Bem.: BF₁₀ = Bayes Faktor

Nach Schönbrodt und Wagenmakers (2017) können Werte unter 0,33 als substanzielle Belege für die Annahme der *H*₀ gelten. Liegen die Werte zwischen 0,33 und 1 werden diese als anekdotische Belege aufgefasst. Für die hier zu testende Hypothese, unterscheiden sich Lehrkräfte und Didaktiker*innen signifikant in der Beurteilung derselben Erklärungen, bedeutet das: In allen sieben Fällen kann angenommen werden, dass die Gültigkeit der *H*₀ wahrscheinlicher ist als die der *H*₁. In drei Fällen liegen substanzielle (TA_a; TL_oa; TL_a), in zwei Fällen nahezu substanzielle Belege vor (NW_o; NW_oa). Die Ergebnisse der Erklärungen TA_oa und TL_a können als anekdotische Belege gezählt werden. Demnach kann *Hypothese 1.1* unterstützt werden, die zwischen den Gruppen der Lehrkräfte und der Didaktiker*innen keine signifikanten Unterschiede in den Globalbewertungen der Erklärqualität erwartet. Dies stützt die Annahme, dass beide Gruppen gemeinsam betrachtet werden können.

Neben dieser Variablen werden außerdem die durchschnittlichen Beurteilungen der Lehrkräfte und Didaktiker*innen zu den jeweiligen Qualitätsmerkmalen über alle sieben Videos hinweg herangezogen.

Tabelle 5.2 Deskriptive Werte der Beurteilungen zu Merkmalen guten Erklärens sowie der Kontrollvariable Persönlichkeitswirkung

Merkmal	Persönlichkeitswirkung	Sprech- & Körperausdruck	Sprachliche Verständlichkeit	Adressatenorientierung	Strukturiertheit	Musikbez. Repräsentationen	
Statusgruppe	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	
Lehrkräfte (<i>N</i> _{max} = 41)	2,65 (2,61)	2,25 (2,10)	2,96 (3,02)	3,33 (3,36)	2,29 (2,26)	1,73 (1,85)	15,22
Didaktiker*innen (<i>N</i> _{max} = 21)	2,38 (2,43)	1,80 (1,95)	2,88 (2,81)	3,16 (3,13)	2,08 (2,10)	1,85 (1,72)	14,15
	5,04	4,05	5,84	6,49	4,37	3,58	29,37

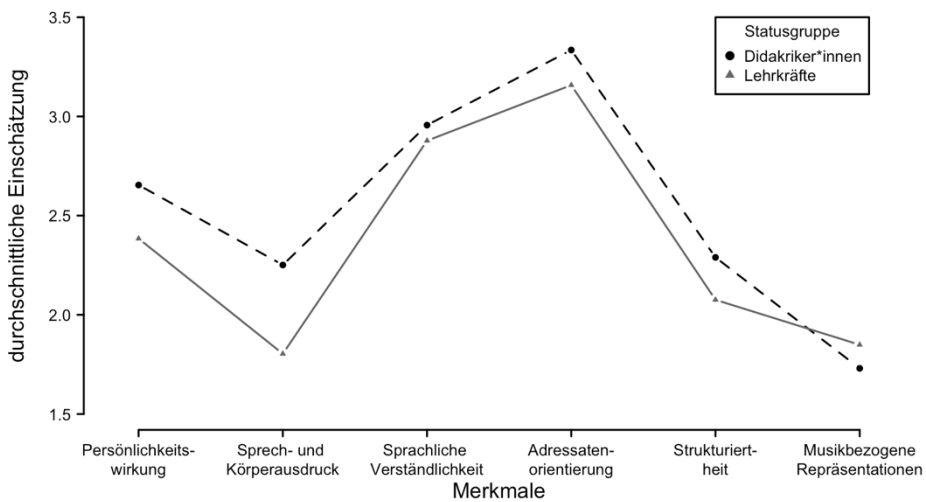
Bem.: *M*: arithm. Mittel; Fett-markiert: Durchschnittliche Werte über alle Videos hinweg; in Klammer: zu erwartender Wert.

Tabelle 5.2 zeigt deskriptive Werte zur Kontrollvariable *Persönlichkeitswirkung* sowie den verschiedenen Merkmalen unterrichtlichen Erklärens beider Gruppen. Es zeigt sich,

dass weder die arithmetischen Mittelwerte mit Ausnahme des Merkmals *Sprech- und Körperausdruck* noch die Standardabweichungen große Unterschiede aufweisen. Auch die Summe der arithmetischen Mittel über alle Merkmale hinweg liegt pro Gruppe mit 15,22 für die Lehrkräfte und 14,15 für die Studierenden nahe beieinander.

Die grafische Darstellung in Form von Profilplots (siehe Abb. 5.1) mit den durchschnittlichen Werten pro Merkmal beider Gruppen stützt die Beobachtung. Es wird angenommen, dass die zufälligen Abweichungen der einzelnen Werte vom Durchschnitt einer χ^2 -Verteilung folgen.

Abbildung 5.1 Profilplots zu den arithmetischen Mittelwerten der Merkmale unterrichtlichen Erklärens getrennt nach Statusgruppen



Um Unterschiede in den durchschnittlichen Werten zu allen Qualitätsmerkmalen und den beiden Gruppen der Lehrkräfte und Didaktiker*innen auf Signifikanz zu überprüfen, wurde ein χ^2 -Test durchgeführt. Die Werte innerhalb der Zellen von Tabelle 5.2 entsprechen dabei keinen Häufigkeiten, die Durchschnittswerte (fett markiert) werden mit den anhand der Zeilen- und Spaltensummen ermittelten Summenwerten verglichen. Der χ^2 -Test ergab keinen signifikanten Zusammenhang zwischen den Durchschnittswerten der Qualitätsmerkmale und den beiden Statusgruppen ($\chi^2(5) = 0,0454, p = 1$). Beide Variablen, die Zugehörigkeit zur Statusgruppe sowie die Beurteilung von Qualitätsmerkmalen unterrichtlichen Erklärens, korrelieren kaum miteinander und sind somit voneinander unabhängig. Das bedeutet, dass auch anhand der durchschnittlichen Beurteilungen der einzelnen Qualitätsmerkmale keine Indizien für eine gesonderte Betrachtung beider Gruppen sprechen und eine gemeinsame Gruppe angenommen werden kann. Damit wird *Hypothese 1.2* unterstützt, die zwischen Lehrkräften und Didaktiker*innen keine signifikanten

Unterschiede in den Beurteilungen zu ausgewählten Qualitätsmerkmalen unterrichtlichen Erklärens erwartet.

Für die weiteren Analysen wird die Gesamtstichprobe daher immer in drei Untergruppen aufgeteilt, die sich aus (1) Schüler*innen (2) Studierenden und (3) Lehrenden (Lehrkräfte und Didaktiker*innen) zusammensetzt. Gleichzeitig bedeutet dies eine Limitation hinsichtlich differenzierter Aussagen aufseiten der Erklärenden, die jedoch mit Blick auf robustere Analysen mit den Stichprobengrößen der verbleibenden drei Gruppen in Kauf genommen wird. Eine Aufteilung in vier Gruppen sollte für anschließende Analysen jedoch nicht ausgeschlossen werden.

5.2 Gütekriterien

5.2.1 Psychometrische Gütekriterien der geschlossenen Fragebogenitems

Objektivität

In der vorliegenden Studie wurde eine möglichst objektive Durchführung gewährleistet, indem der Onlinefragebogen von standardisierten Instruktionen eingeleitet wurde, nach denen die Bearbeitung erfolgte. Alle Studienteilnehmer*innen hatten dieselben Rahmenbedingungen: Es gab keine Zeitbegrenzung und die Items wurden für alle Gruppen in derselben Reihenfolge gestellt. Die Gruppe der Schüler*innen, die alle Fragebögen vor Ort im jeweiligen Computerraum ihrer Schule bearbeitete, erhielt zudem weitere mündliche Instruktionen¹¹⁴ zum Ablauf des Fragebogens vom Testleiter vor Ort. Auch gab es für die Schüler*innen einen Katalog (siehe *Anhang C2*) mit standardisierten Paraphrasierungen zu allen fachübergreifend gestellten Fragebogenitems der Skalen *Adressatenorientierung*, *Strukturiertheit*, *Sprachliche Verständlichkeit*, *Sprech- und Körperausdruck* sowie der Kontrollvariablen zur *Persönlichkeitswirkung*.¹¹⁵ Dieser kam in vier Fällen zum Einsatz, bei denen jeweils das Item „Die Erklärung hatte einen roten Faden“ (Strukturiertheit) nicht verstanden wurde; weitere Fragen wurden nicht gestellt. Ein Großteil der Items wurde mittels geschlossener Antwortformate gestellt, weshalb eine objektive Auswertung der quantitativen Daten gewährleistet ist. In Kapitel 4.9.1 wurde die Datenaufbereitung sowie der Umgang mit fehlenden Werten dokumentiert; außerdem stehen alle Daten des Datensatzes sowie des R-Codes für wiederholende Auswertungen zur Verfügung (siehe *Anhang H*). Da die vorliegende Studie in ihren Grundzügen explorativ angelegt ist und unterschiedliche Perspektiven trianguliert, liegen keine Normen vor, die für eine Interpretation herangezogen werden können. Eine objektive Interpretation der Daten ist damit deutlich eingeschränkt.

¹¹⁴ Diese folgen einer standardisierten Vorlage, siehe *Anhang C1*.

¹¹⁵ Dies erschien notwendig, da in den beteiligten Fächern des Projekts FALKE-q sowohl Schüler*innen der Primar- als auch der Sekundarstufen I und II beteiligt waren. Gerade bei jüngeren Jahrgangsstufen wurde an einzelnen Stellen mit Verständnisschwierigkeiten gerechnet.

Reliabilität

Für die innere Konsistenz der einzelnen Skalen werden Reliabilitätswerte angegeben. Als Kennwert für die Reliabilitäten dient Cronbachs Alpha (z. B. Bühner, 2021, Kap. 4; Field, 2013, Kap. 17.9.1). Zudem werden *part-whole*-korrigierte Itemtrennschärfen (r_{it}) berichtet. Die Tabellen (siehe Tab. 5.3, Tab 5.4 und Tab. 5.5) geben die gemittelte Itemtrennschärfe sowie die dazugehörige Spannweite der Einzelwerte an. Für eine detailliertere Übersicht sind alle Trennschärfen der verwendeten Items nach Gruppen sortiert in den *Anhängen E1, E2* und *E3* zu finden. Als statistische Orientierung für die Interpretation dienen Richtwerte nach Field (2013, S. 675–681). Im Folgenden werden für alle drei Statusgruppen getrennt sowohl die Cronbachs Alpha Werte als auch die durchschnittlichen Itemtrennschärfen sowie deren Spannweite pro Skala und Video berichtet. Die Reihenfolge der Videos entspricht den bisherigen Darstellungen und ist nach Themen zusammengefasst.

Tabelle 5.3 Interne Konsistenzen der erhobenen Skalen für **Schüler*innen (N = 148)** (Cronbachs α und *part-whole*-korrigierte mittlere Itemtrennschärfen r_{it} sowie deren Range)

Video	NW_o α	NW_oa α	TA_a α	TA_oa α	TL_o α	TL_a α	TL_oa α
Merkmal	r_{it} : M (SD) range	r_{it} : M (SD) range	r_{it} : M (SD) range	r_{it} : M (SD) range	r_{it} : M (SD) range	r_{it} : M (SD) range	r_{it} : M (SD) range
Sprech- und Körperausdruck (7 Items)	0,78 0,51 (0,12) 0,34-0,64	0,84 0,60 (0,07) 0,49-0,70	0,84 0,59 (0,13) 0,43-0,79	0,93 0,78 (0,08) 0,61-0,86	0,91 0,73 (0,06) 0,60-0,77	0,93 0,77 (0,05) 0,66-0,82	0,86 0,62 (0,06) 0,55-0,71
Sprachliche Verständlichkeit (3 Items)	0,40 0,34 (0,34) -0,05-0,61	0,45 0,37 (0,25) 0,00-0,61	0,48 0,36 (0,25) 0,08-0,52	0,41 0,32 (0,28) 0,00-0,50	0,35 0,27 (0,24) -0,01-0,45	0,42 0,33 (0,25) 0,04-0,50	0,32 0,28 (0,34) -0,10-0,52
Adressatenorientierung (3 Items)	0,42 0,22 (0,22) 0,22-0,48	0,56 0,35 (0,16) 0,12-0,43	0,47 0,27 (0,16) 0,21-0,49	0,52 0,34 (0,23) 0,28-0,60	0,68 0,47 (0,20) 0,15-0,60	0,55 0,34 (0,19) 0,30-0,57	0,58 0,38 (0,23) 0,23-0,54
Strukturiertheit (3 Items)	0,75 0,56 (0,09) 0,46-0,65	0,74 0,56 (0,06) 0,52-0,62	0,82 0,66 (0,07) 0,58-0,73	0,80 0,63 (0,12) 0,50-0,73	0,75 0,58 (0,03) 0,55-0,61	0,79 0,62 (0,06) 0,55-0,68	0,79 0,62 (0,05) 0,57-0,67
Musikbez. Repräsentationen (kurz) (4 Items)	0,71 0,49 (0,05) 0,45-0,54	0,66 0,43 (0,12) 0,32-0,57	0,83 0,65 (0,07) 0,56-0,72	0,80 0,62 (0,08) 0,59-0,73	0,69 0,46 (0,12) 0,35-0,61	0,76 0,55 (0,07) 0,49-0,63	0,82 0,64 (0,04) 0,60-0,69
Musikbez. Repräsentationen (lang)	0,75 0,51 (0,04) 0,47-0,58	0,83 0,55 (0,09) 0,39-0,67	0,83 0,65 (0,07) 0,56-0,72	0,86 0,61 (0,07) 0,50-0,72	0,73 0,50 (0,14) 0,34-0,68	0,81 0,60 (0,09) 0,46-0,69	0,86 0,60 (0,06) 0,55-0,72
Anzahl Items	5	8	4	8	5	5	8
Persönlichkeitswirkung (3 Items)	0,68 0,49 (0,08) 0,40-0,55	0,87 0,75 (0,05) 0,70-0,80	0,71 0,53 (0,08) 0,47-0,62	0,90 0,79 (0,04) 0,76-0,83	0,92 0,83 (0,01) 0,82-0,84	0,92 0,84 (0,03) 0,81-0,86	0,71 0,52 (0,11) 0,41-0,63

Bem.: M: arithm. Mittel; SD: Standardabweichung; range: Minimum-Maximum.

Tabelle 5.3 zeigt die Werte für die Gruppe der Schüler*innen. Dabei liegen für die Skalen *Sprech- und Körperausdruck*, *Strukturiertheit*¹¹⁶, *Einsatz musikbezogener Repräsentationen* (sowohl in der Kurz- als auch der Langversion) sowie die Kontrollvariable *Persönlichkeitswirkung* neben wenigen Ausnahmen mittlere bis hohe Cronbachs Alpha Werte vor. Die dazugehörigen Werte der Itemtrennschärfe liegen durchweg im mittleren bis hohen Bereich, weshalb alle genannten Skalenreliabilitäten vor dem Hintergrund der Neukonstruktion sowie den geringen Skalenlängen als zufriedenstellend eingestuft werden können. Für die Skalen *Sprachliche Verständlichkeit*¹¹⁷ und *Adressatenorientierung*¹¹⁸ liegen die Werte für Cronbachs Alpha über alle Videos hinweg im niedrigen bis sehr niedrigen Bereich. Unter Einbezug der Itemtrennschärfen zeigt sich, dass bei der Skala *Adressatenorientierung* jeweils vor allem das Item zu Lebensweltbezug (*Die Inhalte der Erklärung hatten etwas mit meinem Leben zu tun*) niedrige Itemtrennschärfen aufweist (siehe *Anhang E1*). Es wird vermutet, dass die fachübergreifende Formulierung mit Blick auf die Inhalte im Fach Musik zu Missverständnissen und Fehldeutungen führen kann. Wie aus den theoretischen Überlegungen (siehe Kap. 2.3.2) hervorgeht, stellt das Konzept Lebensweltbezug jedoch eine wichtige Facette der *Adressatenorientierung* dar, weshalb das Item in der Skala belassen wird, um Vergleiche mit den Werten der anderen Statusgruppen zu ermöglichen.

Bei der Skala *Sprachliche Verständlichkeit* korreliert insbesondere das Item *Der Lehrer hat sich gut verständlich ausgedrückt* nicht mit den anderen beiden Skalenitems und nimmt Itemtrennschärfe-Werte im Bereich von $-0,10 < r_{it} < 0,08$ an. Um die Vergleichbarkeit mit den Skalenwerten der Studierenden und Lehrenden, die deutlich höhere Reliabilitäten zeigen, zu gewährleisten, wird dieses Item jedoch auch bei den Schüler*innen in der Skala belassen. Allerdings führt dies bei der Analyse der Ergebnisse zu weitreichenden Limitationen, die insbesondere deshalb zu berücksichtigen sind, da bei dieser Skala bereits das Item *Manchmal hat der Lehrer extra gesagt, dass etwas sehr wichtig ist* aufgrund negativer Korrelationen in allen drei Gruppen eliminiert werden musste. Die Skala *Sprachliche Verständlichkeit* hat auch in den anderen Gruppen niedrige Reliabilitätswerte.

Aus Tabelle 5.4 gehen die Werte für die Gruppe der Studierenden hervor. Dabei zeigen sich über alle Videos hinweg mittlere bis hohe Werte von Cronbachs Alpha für die

116 Aufgrund negativer Korrelationen des Items *Der Lehrer hat zu viele Einzelheiten erklärt* (St_4, siehe Tab. 4.6) in den Erklärungen NW_o, NW_oa, TA_a, TA_oa, TL_a und TL_oa wurde das Item bei der Gruppe der Schüler*innen aus der Skala entfernt. Um Vergleichbarkeit mit den anderen Statusgruppen zu erzielen, wird das Item auch dort ausgeschlossen – zudem liegen auch dort negative Korrelationen dieses Items vor (siehe *Anhang E*).

117 Aufgrund negativer Korrelationen des Items *Manchmal hat der Lehrer extra gesagt, dass etwas sehr wichtig ist* (SV_2, siehe Tab. 4.5) in allen Erklärungen wurde das Item bei der Gruppe der Schüler*innen aus der Skala entfernt. Um Vergleichbarkeit mit den anderen Statusgruppen zu erzielen, wird das Item auch dort ausgeschlossen – zudem liegen auch dort negative Korrelationen dieses Items vor (siehe *Anhang E*).

118 Aufgrund negativer Korrelationen des Items *Die Erklärung war für mich zu schwierig* (Ad_1, siehe Tab. 4.7) in allen Erklärungen wurde das Item bei der Gruppe der Schüler*innen aus der Skala entfernt. Da die Skala bei Studierenden und Lehrenden umfassender ist und mit einem Item zur didaktischen Reduktion der Erklärung ohnehin eine andere Itemzahl hat, erfolgt dort kein Ausschluss des Items (siehe *Anhang E*).

Skalen *Sprech- und Körperausdruck*, *Adressatenorientierung*¹¹⁹, *Strukturiertheit* sowie *Persönlichkeitswirkung*. Etwas niedrigere Werte liegen für *Sprachliche Verständlichkeit* vor, während die beiden Varianten der Skala *Einsatz musikbezogener Repräsentationen* (kurz und lang) generell niedrige Werte von Cronbachs Alpha vorweisen. Diese können auch mit der niedrigen Zahl an Studierenden einhergehen, da die Stichprobe ($N = 61$) für solche Analysen klein ausfällt.

Tabelle 5.4 Interne Konsistenzen der erhobenen Skalen für **Studierende ($N = 61$)** (Cronbachs α und *part-whole*-korrigierte mittlere Itemtrennschärfen r_{it} sowie deren Range)

Video Merkmal	NW_o	NW_oa	TA_a	TA_oa	TL_o	TL_a	TL_oa
	α $r_{it}: M (SD)$ range	α $r_{it}: M (SD)$ range	α $r_{it}: M (SD)$ range	α $r_{it}: M (SD)$ range	α $r_{it}: M (SD)$ range	α $r_{it}: M (SD)$ range	α $r_{it}: M (SD)$ range
Sprech- und Körperausdruck (7 Items)	0,75 0,47 (0,16) 0,19-0,72	0,77 0,50 (0,17) 0,27-0,72	0,82 0,56 (0,14) 0,36-0,75	0,86 0,62 (0,08) 0,50-0,76	0,88 0,67 (0,10) 0,45-0,75	0,81 0,54 (0,08) 0,42-0,69	0,77 0,49 (0,17) 0,14-0,72
Sprachliche Verständlichkeit (3 Items)	0,76 0,60 (0,10) 0,50-0,67	0,65 0,47 (0,09) 0,38-0,57	0,74 0,57 (0,08) 0,48-0,62	0,63 0,45 (0,10) 0,34-0,53	0,64 0,51 (0,23) 0,25-0,68	0,74 0,56 (0,04) 0,53-0,60	0,60 0,43 (0,10) 0,31-0,49
Adressatenorientierung (5 Items)	0,81 0,59 (0,19) 0,35-0,81	0,80 0,57 (0,12) 0,35-0,67	0,83 0,64 (0,22) 0,27-0,82	0,84 0,64 (0,15) 0,39-0,75	0,83 0,63 (0,11) 0,47-0,74	0,85 0,66 (0,13) 0,51-0,76	0,79 0,56 (0,15) 0,35-0,68
Strukturiertheit (3 Items)	0,81 0,66 (0,08) 0,61-0,75	0,80 0,65 (0,09) 0,56-0,74	0,84 0,71 (0,05) 0,66-0,76	0,85 0,72 (0,09) 0,65-0,83	0,75 0,58 (0,12) 0,45-0,70	0,84 0,71 (0,04) 0,68-0,75	0,84 0,70 (0,07) 0,64-0,78
Musikbez. Repräsentationen (kurz) (4 Items)	0,46 0,26 (0,05) 0,20-0,30	0,55 0,33 (0,08) 0,24-0,42	0,72 0,53 (0,18) 0,46-0,68	0,72 0,50 (0,19) 0,29-0,62	0,35 0,14 (0,11) 0,02-0,26	0,56 0,31 (0,19) 0,07-0,53	0,62 0,40 (0,10) 0,26-0,46
Musikbez. Repräsentationen (lang) (3 Items)	0,39 0,17 (0,09) 0,04-0,26	0,63 0,32 (0,12) 0,18-0,50	0,72 0,53 (0,18) 0,31-0,68	0,75 0,44 (0,14) 0,28-0,70	0,46 0,26 (0,19) 0,02-0,46	0,59 0,34 (0,13) 0,13-0,46	0,63 0,31 (0,12) 0,05-0,48
Anzahl Items	5	8	4	8	5	5	8
Persönlichkeitswirkung (3 Items)	0,77 0,61 (0,06) 0,54-0,62	0,41 0,22 (0,18) 0,10-0,42	0,82 0,67 (0,09) 0,60-0,78	0,75 0,58 (0,04) 0,53-0,62	0,72 0,54 (0,09) 0,44-0,60	0,69 0,50 (0,03) 0,48-0,54	0,80 0,64 (0,12) 0,50-0,71

Bem.: M : arithm. Mittel; SD : Standardabweichung; *range*: Minimum-Maximum.

Ein klareres Bild liefern auch hier die Werte zu den Itemtrennschärfen. Diese zeigen bei der Skala *Sprech- und Körperausdruck* einzelne Ausreißerwerte von jeweils unterschiedlichen Items bei den Erklärungen NW_o, NW_oa und TL_oa (siehe *Anhang E*). Da jeweils drei unterschiedliche Items bei den drei Erklärungen zu niedrigen Werten führen, werden alle Items in der Skala belassen. Die niedrigeren Werte der Skala *Sprachliche*

¹¹⁹ Die Anzahl an Items variiert bei den Gruppen der Studierenden und Lehrenden im Vergleich zur Gruppe der Schüler*innen, da das Item zur didaktischen Reduktion der Erklärungen von der Gruppe der Schüler*innen aufgrund deren fehlender Expertise nicht adäquat beantwortet werden kann.

Verständlichkeit sind insbesondere auf das Item *Der Lehrer hat sich gut verständlich ausgedrückt* zurückzuführen. Ähnlich wie bei der Gruppe der Schüler*innen bleibt das Item auch hier in der Skala, um die inhaltliche Breite des Konstrukts beizubehalten. Für die Skala *Adressatenorientierung* lassen sich die jeweils niedrigsten Werte der Itemtrennschärfe auf das Item zu Lebensweltbezug zurückführen. Obwohl sich die Formulierung dieses Items etwas von der für die Schüler*innen unterscheidet, damit die Perspektive auf Erklären sowie das dahinterliegende didaktische Verständnis des Begriffs Lebenswelt berücksichtigt werden kann, und die Skala ein weiteres Item zur didaktischen Reduktion abfragt, lassen sich auch hier die jeweils niedrigen Werte diesem Item zuordnen. Im Gegensatz dazu sind bei beiden Versionen zur Skala *Einsatz musikbezogener Repräsentationen* unterschiedliche Items für niedrige Itemtrennschärfe-Werte verantwortlich. Um die breite Erfassung beider Skalen beizubehalten, werden aus inhaltlichen Gründen keine Items eliminiert. Allerdings ist dies als Limitation bei der Darstellung der Ergebnisse zu berücksichtigen.

Schließlich berichtet Tabelle 5.5 die Reliabilitätswerte für alle Skalen der Lehrenden. Dabei zeigen sich in den Skalen *Sprech- und Körperausdruck*, *Adressatenorientierung* und *Strukturiertheit* durchweg mittlere bis hohe Werte von Cronbachs Alpha.

Tabelle 5.5 Interne Konsistenzen der erhobenen Skalen für **Lehrende (N = 62)** (Cronbachs α und *part-whole*-korrigierte mittlere Itemtrennschärfen r_{it} sowie deren Range)

Video	NW_o	NW_oa	TA_a	TA_oa	TL_o	TL_a	TL_oa
	α	α	α	α	α	α	α
Merkmal	$r_{it}: M (SD)$	$r_{it}: M (SD)$	$r_{it}: M (SD)$	$r_{it}: M (SD)$	$r_{it}: M (SD)$	$r_{it}: M (SD)$	$r_{it}: M (SD)$
	range	range	range	range	range	range	range
Sprech- und Körperausdruck (7 Items)	0,86 0,62 (0,12) 0,47-0,77	0,88 0,67 (0,18) 0,31-0,82	0,67 0,38 (0,13) 0,24-0,63	0,90 0,72 (0,09) 0,59-0,85	0,83 0,58 (0,08) 0,49-0,69	0,95 0,83 (0,07) 0,70-0,92	0,84 0,58 (0,04) 0,54-0,63
Sprachliche Verständlichkeit (3 Items)	0,48 0,32 (0,16) 0,17-0,49	0,81 0,66 (0,02) 0,65-0,68	0,70 0,52 (0,03) 0,50-0,55	0,57 0,40 (0,12) 0,26-0,47	0,55 0,37 (0,10) 0,31-0,48	0,68 0,49 (0,06) 0,43-0,55	0,61 0,41 (0,08) 0,32-0,48
Adressatenorientierung (5 Items)	0,88 0,70 (0,11) 0,51-0,80	0,90 0,74 (0,13) 0,52-0,87	0,89 0,72 (0,13) 0,49-0,80	0,86 0,67 (0,14) 0,46-0,83	0,84 0,63 (0,13) 0,44-0,72	0,87 0,69 (0,11) 0,54-0,83	0,86 0,67 (0,12) 0,56-0,83
Strukturiertheit (3 Items)	0,85 0,71 (0,06) 0,64-0,77	0,94 0,87 (0,05) 0,82-0,92	0,71 0,51 (0,21) 0,29-0,72	0,79 0,62 (0,10) 0,51-0,72	0,83 0,69 (0,13) 0,55-0,80	0,89 0,78 (0,10) 0,69-0,89	0,79 0,62 (0,07) 0,55-0,67
Musikbez. Repräsentationen (kurz) (4 Items)	0,72 0,49 (0,15) 0,30-0,65	0,65 0,42 (0,11) 0,35-0,58	0,74 0,54 (0,22) 0,26-0,80	0,73 0,52 (0,08) 0,41-0,59	0,65 0,43 (0,18) 0,17-0,58	0,74 0,48 (0,16) 0,27-0,63	0,70 0,45 (0,24) 0,22-0,71
Musikbez. Repräsentationen (lang)	0,68 0,39 (0,13) 0,22-0,54	0,77 0,47 (0,13) 0,33-0,73	0,74 0,54 (0,22) 0,26-0,80	0,86 0,62 (0,15) 0,42-0,77	0,65 0,42 (0,22) 0,10-0,72	0,73 0,47 (0,08) 0,38-0,57	0,80 0,50 (0,19) 0,12-0,66
Anzahl Items	5	8	4	8	5	5	8
Persönlichkeitswirkung (3 Items)	0,57 0,39 (0,14) 0,24-0,51	0,74 0,56 (0,06) 0,52-0,63	0,58 0,41 (0,13) 0,30-0,55	0,81 0,67 (0,07) 0,59-0,71	0,68 0,50 (0,05) 0,44-0,55	0,86 0,73 (0,03) 0,70-0,77	0,52 0,33 (0,20) 0,14-0,54

Bem.: *M*: arithm. Mittel; *SD*: Standardabweichung; *range*: Minimum-Maximum.

Die beiden niedrigen Werte der Skalen *Sprech- und Körperausdruck* sowie *Strukturiertheit* lassen sich bei der Erklärung TA_a anhand der Itemtrennschärfe-Werte auf jeweils eine einzelne Ausnahme zurückführen (siehe *Anhang E3*), die jedoch nicht den Ausschluss des Items rechtfertigt.

Bei *Sprachliche Verständlichkeit* liegen auch in der Gruppe der Lehrenden niedrigere Werte vor, die sich beim Blick auf die Analyse der Itemtrennschärfen ebenfalls anhand des Items *Der Lehrer hat sich gut verständlich ausgedrückt* erklären. Auch hier erfolgt kein Ausschluss, sodass die Breite des Konstrukts weiterhin abgebildet ist. Für die Skalen zum *Einsatz musikbezogener Repräsentationen* zeigen einzelne Items bei verschiedenen Erklärungen niedrige Werte. Um die Vergleichbarkeit mit den beiden anderen Gruppen zu gewährleisten und die Breite des Konstrukts abzudecken, werden auch hier keine Items von der Skala ausgeschlossen. Dies gilt auch für die Skala *Persönlichkeitswirkung*, die sowohl niedrige bis mittlere Werte für Cronbachs Alpha als teilweise auch niedrige Itemtrennschärfen aufweist. Diese resultieren für die Erklärungen NW_o und TL_a jeweils für das Item *Der Lehrer war natürlich*. Darüber hinaus kann – analog zu den Studierenden – auch bei den Werten der Lehrenden die niedrige Stichprobenzahl ($N = 62$) zu den niedrigen Werten führen.

Insgesamt zeigt sich über die Reliabilitätsanalysen der drei Gruppen hinweg, dass die Skalen *Sprech- und Körperausdruck*, *Strukturiertheit*, *Einsatz musikbezogener Repräsentationen* (kurz und lang) sowie die Kontrollvariablen zur *Persönlichkeitswirkung* mit wenigen Ausnahmen vor dem Hintergrund der Neukonstruktion als zufriedenstellend eingestuft werden können. Die Kurz- sowie Langversion der Skala *Einsatz musikbezogener Repräsentationen* zeigt insbesondere für die Gruppe der Schüler*innen akzeptable bis gute Werte. Trotz einiger Ausnahmen niedriger Werte für Studierende und Lehrende erfolgt kein Itemausschluss, da über alle Erklärungen hinweg keine systematisch niedrigen Werte einzelner Items vorliegen. Zudem bleiben die Skalenwerte zwischen den Gruppen aufgrund gleicher Skalenlänge vergleichbar.

Die Skala *Adressatenorientierung* differiert zwischen den Versionen beider Perspektiven (Schüler*innen auf der einen sowie Studierende und Lehrende auf der anderen Seite) in Skalenlänge sowie der Formulierung zum Konzept Lebensweltbezug. Daher müssen diese getrennt betrachtet werden. In beiden Skalen weist insbesondere das Item zu Lebensweltbezug niedrigere Werte für Cronbachs Alpha und Itemtrennschärfen auf. Mit dem Ziel einer breiten inhaltlichen Abbildung der Skala wird kein Ausschluss des Items vorgenommen, obwohl an manchen Stellen ein niedriger Wert von r_{it} vorliegt.

Problematisch erweist sich insbesondere die Skala *Sprachliche Verständlichkeit*. Neben der Eliminierung des Items *Manchmal hat der Lehrer extra gesagt, dass etwas sehr wichtig ist* aufgrund negativer Korrelationen in allen Gruppen (siehe *Anhang E*) liegen auch für das Item *Der Lehrer hat sich gut verständlich ausgedrückt* niedrige Itemtrennschärfen vor. Obwohl gerade die Werte bei den Schüler*innen zum Ausschluss des Items führen müssten, bleibt das Item in allen Gruppen in der Skala. Damit soll zum einen die Vergleichbarkeit zwischen den Gruppen gewährleistet bleiben, zum anderen kann so die inhaltliche Dimensionierung der Wort-, Satz- und Textebene (die Erklärung wird aus

linguistischer Sicht als Text verstanden) beibehalten werden. Bei der Darstellung und Diskussion der Ergebnisse werden diese Limitationen berücksichtigt.

Validität

Zur Beurteilung des vorliegenden Fragebogens werden Inhaltsvalidität und Konstruktvalidität herangezogen.¹²⁰ Die Inhaltsvalidität eines Fragebogens ergibt sich primär aus fachlichen Gründen (siehe Bühner, 2021, S. 601). So wurde hinsichtlich der ökologischen Validität der Erklärsituationen abgefragt, ob diese im Musikunterricht aus Sicht von Studierenden und Lehrenden vorkommen (Kv_4; siehe Tab. 4.10).¹²¹ Über alle sieben Videos ergibt sich ein Durchschnittswert von 2,25, der in etwa der Angabe *stimme zu* entspricht. Vor dem Hintergrund einer sechsstufigen Zustimmungsskala bedeutet dies, dass grundsätzlich von ökologischer Validität der Stimuli ausgegangen werden kann.¹²² Interessant ist in diesem Zusammenhang der Blick auf die einzelnen Videos (siehe Tab. 5.6).

Tabelle 5.6 Arithmetisches Mittel und Standardabweichung von Item Kv_4

Itemformulierung	NW_o <i>M (SD)</i>	NW_oa <i>M (SD)</i>	TA_a <i>M (SD)</i>	TA_oa <i>M (SD)</i>	TL_o <i>M (SD)</i>	TL_a <i>M (SD)</i>	TL_oa <i>M (SD)</i>	Gesamt <i>M (SD)</i>
Erklärsituationen wie die im Video gibt es im Musikunterricht.	2,13 (1,14)	2,23 (1,21)	2,48 (1,19)	1,19 (1,12)	2,08 (1,05)	2,66 (1,44)	1,97 (1,09)	2,25 (1,18)

Bem.: *M*: arithm. Mittel; *SD*: Standardabweichung.

Aus Tabelle 5.6 geht hervor, dass insbesondere die Erklärungen die beiden schlechtesten Werte haben, in denen eine optische Repräsentation fehlt (TA_a; TL_a). Dies kann als erster Hinweis darauf dienen, dass solche Erklärungen im Musikunterricht tendenziell seltener vorkommen. Der Erstellungsprozess von Erklärvideos und Skalen wurde einerseits durch Rückmeldungen von Expert*innen aus Schule und Lehrkräftebildung begleitet, auf deren Basis Modifikationen vorgenommen wurden. Andererseits konnten durch Ergebnisse kognitiver Interviews mit Vertreter*innen aller beteiligter Statusgruppen mittels der *Probing*-Technik auch inhaltliche Aspekte überprüft werden (siehe Frei, 2019; siehe Kap. 4.7). Zudem ermöglicht die Kombination aus quantitativen und qualitativen Daten eine breite inhaltliche Erfassung des Konstrukts *Erklären*, sodass blinde Flecken in der Erhebung möglichst ausgeschlossen werden können.

Die Konstruktvalidität kann anhand von zwei Aspekten überprüft werden. Zum einen kann mithilfe der offenen Antworten von Studierenden und Lehrenden analysiert werden,

¹²⁰ Die strukturelle Validität kann mit einer konfirmatorischen Faktorenanalyse überprüft werden. In der vorliegenden Arbeit wird auf diese Analyse verzichtet, da hierzu keine Forschungsfrage vorliegt.

¹²¹ In Kapitel 4.7 wurde als ein Ergebnis der Pilotierung begründet, warum dieses Item den Schüler*innen nicht gestellt wurde.

¹²² Aussagen hinsichtlich realen Musikunterrichts müssen dennoch die Labvorsituation berücksichtigen, in der die Erklärungen gegeben wurden.

mit welchen Merkmalen sie die Qualität des Erklärens begründen. Da zu diesem Zeitpunkt der Erhebung noch keine Skalen zu den einzelnen Merkmalen abgefragt werden, können die getroffenen Aussagen als Validierung der ausgewählten Merkmale guten Erklärens dienen. Zum anderen sollen die Ergebnisse der Regressionsanalysen pro Statusgruppe den jeweiligen Einfluss der einzelnen Merkmale abbilden, indem inferenzstatistisch signifikante Zusammenhänge zwischen Erklärqualität und Qualitätsmerkmalen aufgezeigt werden. Da die Annahme einer perspektivenspezifischen Validität besteht, werden die Ergebnisse aller drei Gruppen zur Einschätzung der Konstruktvalidität berücksichtigt. Das bedeutet, dass auch von Konstruktvalidität gesprochen wird, wenn signifikante Zusammenhänge zwischen einzelnen Merkmalen und der Bewertung der Erklärqualität nur in einer der drei Gruppen vorliegen.

5.2.2 Psychometrische Gütekriterien des Wissenstests

Objektivität

Analog zur Darstellung der psychometrischen Gütekriterien des Fragebogens erfolgt eine knappe Einordnung hinsichtlich des Wissenstests. Standardisierte Instruktionen sowie dieselben Rahmenbedingungen für alle Schüler*innen stellen eine objektive Durchführung des Wissenstests sicher. Eine objektive Auswertung kann für die geschlossenen Items gewährleistet werden. Hinsichtlich der Aufgaben, die von den Schüler*innen in einem Textfeld zu beantworten waren, wurde ein Bewertungsmanual (siehe *Anhang B4*) erstellt und einer zweiten Person¹²³ zum Codieren vorgelegt. Bei den Antworten, die beide Personen unterschiedlich codiert haben, konnte im Sinne des konsensuellen Codierens (siehe Kuckartz, 2016, S. 211–212) in allen Fällen eine Einigung erzielt werden. Es liegen Daten von 142 Lernenden zu allen Testaufgaben vor.

Reliabilität

Die Zuverlässigkeit und Genauigkeit der Messung mithilfe des Wissenstests wird wie in Kapitel 5.2.1 anhand zweier Kenngrößen angegeben. Zum einen werden Reliabilitätswerte zur inneren Konsistenz der drei themenbezogenen Kurz-Wissenstests anhand des Konsistenzkoeffizienten Cronbachs Alpha angegeben. Zum anderen dienen *part-whole*-korrigierte Itemtrennschärfen (r_1) dazu, differenzierte Aussagen auf Itemebene zu tätigen. Die statistische Einordnung erfolgt anhand der Werte, die in Tabelle 5.3 (siehe Kap. 5.2.1) dargelegt wurden.

In Kapitel 4.4.2 wurde erläutert, weshalb für den vorliegenden Test geringere Werte von Cronbachs Alpha erwartet werden. Tabelle 5.7 stellt die Werte der drei Kurz-Wissenstests pro Thema zum Prä-Zeitpunkt dar. Dabei zeigt sich erwartungsgemäß, dass die

¹²³ Es handelt sich dabei um die Person, die auch die zweite Codierung der offenen Antworten des Fragebogens übernommen hat (siehe Kap. 5.2.3).

Werte von Cronbachs Alpha allesamt in einem sehr niedrigen Bereich liegen. Die Itemtrennschärfen liegen insbesondere beim Prä-Wissenstest zum Thema Tonleiter als auch beim Post-Wissenstest zum Thema Notenwerte in einem sehr niedrigen Bereich. Aufgrund dieser Reliabilitätswerte sind die Ergebnisse inferenzstatistischer Analysen (siehe Kap. 5.5) einigen Limitationen unterzogen. Im Vordergrund stehen daher deskriptive Analysen.

Tabelle 5.7 Interne Konsistenzen der drei Wissenstests zu den Themen Notenwerte, Taktarten und Tonleiter für die Gruppe der **Schüler*innen (N = 142)** für den ersten und zweiten Messzeitpunkt (Cronbachs α und *part-whole*-korrigierte mittlere Itemtrennschärfen r_{it} sowie deren Range)

Thema	Notenwerte	Taktarten	Tonleiter
	α r_{it} : M (SD) range	α r_{it} : M (SD) range	α r_{it} : M (SD) range
Messzeitpunkt	0,41	0,44	0,41
	MZP 1 (Prä) 0,24 (0,06) 0,20-0,31	0,27 (0,05) 0,22-0,32	0,24 (0,10) 0,13-0,33
MZP 2 (Post)	0,38	0,55	0,49
	0,21 (0,10) 0,13-0,33	0,37 (0,09) 0,32-0,47	0,31 (0,04) 0,27-0,35

Bem.: M : arithm. Mittel; SD : Standardabweichung; range: Spannweite.

Validität

Bei der Entwicklung des Wissenstests wurden Rückmeldungen von Expert*innen aus Schule und Lehrkräftebildung eingeholt, um die inhaltliche Validität zu steigern. Auch wurden die Aufgaben im Vorfeld mit drei Schüler*innen pilotiert. Anhand der Rückmeldungen konnten Missverständnisse hinsichtlich Instruktion oder Aufgabenstamm reduziert werden (siehe Frei, 2019; siehe Kap. 4.7). Die Konstruktvalidität der drei themenbezogenen Kurz-Wissenstests ist eingeschränkt. Zwar orientieren sich die Aufgaben an curricularen Vorgaben zum Kompetenz- und Wissensstand von Schüler*innen einer sechsten Jahrgangsstufe an bayerischen Gymnasien (Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2017a) und Realschulen (Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2017b), allerdings kann allein aufgrund der geringen Zahl an Aufgaben nicht die komplette inhaltliche Bandbreite der drei Themen abgedeckt werden.

5.2.3 Spezifische Gütekriterien der offenen Fragebogenitems

Die vorliegende Arbeit orientiert sich an den Empfehlungen von Kuckartz (2016), auf Gütekriterien quantitativer Forschung Bezug zu nehmen, diese im Kontext der qualitativen Inhaltsanalyse jedoch „zu reformulieren, teilweise auch neu zu formulieren, und so zu spezifischen Standards zu gelangen“ (S. 203). Dabei erscheint eine engere Anbindung

an Kriterien quantitativer Forschung sinnvoll, da die qualitativen Daten nicht nur im Kontext quantitativer Daten (Globalurteil zur Erklärqualität) erhoben wurden, sondern auch in Bezug darauf ausgewertet werden. Der Autor unterscheidet primär zwischen „interner Studiengüte“ und „externer Studiengüte“ (Kuckartz, 2016, S. 203). Zweite zielt auf die Übertragbarkeit der Ergebnisse und orientiert sich demnach am Produkt der Forschung (siehe auch Strübing et al., 2018). Die interne Studiengüte fokussiert den Forschungsprozess und zeichnet sich durch Kriterien wie Verlässlichkeit, Regelgeleitetheit oder intersubjektive Nachvollziehbarkeit aus (Kuckartz, 2016, S. 203–204). Diese gelten für inhaltsanalytische Auswertungen und weisen Überlappungen mit Kriterien auf, die Mayring (2015, Kap. 7) für die Inhaltsanalyse vorschlägt. Die Güte der qualitativen Daten aus der vorliegenden Studie wird nun anhand ausgewählter Kriterien beider Autoren (Kuckartz, 2016, Kap. 9; Mayring, 2015, Kap. 7 und 2016, Kap. 6) eingeschätzt.

Verfahrensdokumentation

Dem Gütekriterium der transparenten und nachvollziehbaren Dokumentation der Verfahrensweise (siehe Kuckartz, 2016, S. 222; Mayring, 2016, S. 144–145) trägt die Studie Rechnung, indem einerseits der Forschungsstand theoretisch hergeleitet (siehe Kap. 2) und entsprechende Forschungsfragen (siehe Kap. 3.2) gestellt wurden. Andererseits wurde die Triangulation von quantitativen und qualitativen Methoden dargelegt und begründet (Kap. 4.1.2) sowie der Forschungsprozess und das inhaltsanalytische Vorgehen bei der Auswertung der offenen Antworten in Kapitel 4.9.2 ausführlich erläutert. Dazu zählt auch die Verortung innerhalb der inhaltlich strukturierenden qualitativen Inhaltsanalyse nach Kuckartz (2016, Kap. 5), die Entwicklung des Kategoriensystems, die Analyse durch zwei Codierer*innen sowie das Vorgehen bei der Auswahl an Inhalten, die im Ergebnisteil präsentiert werden.

Regelgeleitetheit

Die vorliegende Arbeit kommt dem Kriterium der Regelgeleitetheit nach, indem sie sich dem Vorgehen der inhaltlich strukturierenden qualitativen Inhaltsanalyse nach Kuckartz (2016, Kap. 5) verpflichtet und die dabei vorgeschlagenen Verfahrensschritte weitgehend einhält (siehe Kap. 4.9.2). Die Regeln, nach denen der Textkorpus zu codieren war, sind nachvollziehbar in einem Kategoriensystem dokumentiert (siehe *Anhang F4*). Das Codieren erfolgte schrittweise und war zyklisch angelegt: Abstimmungen zwischen dem Autor der Arbeit und der zweiten Codiererin erfolgten nach jeder der sieben Erklärungen. So konnten spezifische Codierungen bezüglich Erklärthema oder Erklärmodus fokussiert werden; notwendige Adaptionen im Kategoriensystem wurden über den gesamten Prozess hinweg direkt umgesetzt. Im Zusammenspiel mit der detaillierten Verfahrensdokumentation soll auch das Prinzip der Regelgeleitetheit auf den verschiedenen genannten Ebenen des Forschungsprozesses dazu dienen, die intersubjektive Nachvollziehbarkeit der qualitativen Inhaltsanalyse zu erhöhen.

Intercoder-Übereinstimmung

Mayring (2015, S. 124–125) bezeichnet die Intercoder-Reliabilität als spezifisch inhaltsanalytisches Kriterium, das sowohl bei qualitativer wie quantitativer Inhaltsanalyse zum Einsatz kommt. Zudem können damit auch Aussagen zur Objektivität beziehungsweise über die „Unabhängigkeit der Ergebnisse von der untersuchenden Person“ (Mayring, 2015, S. 124) getroffen werden. Kuckartz (2016) nutzt den Begriff der Intercoder-Reliabilität ausschließlich im Kontext der quantitativen Inhaltsanalyse. Für das qualitative Verfahren schlägt er den Terminus „Intercoder-Übereinstimmung“ vor, um damit den „mit dem Anspruch der Replizierbarkeit verknüpften Begriff Reliabilität“ (S. 206) zu vermeiden.

Die vorliegende Studie schließt sich dieser Unterscheidung an, weshalb im Folgenden die Kennwerte zur Intercoder-Übereinstimmung zwischen dem Autor dieser Arbeit (Person 1) und einer zweiten Codiererin (Person 2) angegeben werden. Aufgrund der unterschiedlich ausführlichen Antworten – der Umfang offener Antworten reicht von einzelnen Wörtern bis hin zu mehreren Sätzen –, der insbesondere bei den kurzen Antworten notwendig werdenden Interpretation durch die codierende Person und dem gleichzeitig umfangreichen und differenzierten Kategoriensystem wurde der Wert 70 als Prozentvorgabe für die Übereinstimmung der codierten Segmente zwischen beiden Codierer*innen festgelegt.

Tabelle 5.8 berichtet die Auswertungen für alle Videos und über alle Kategorien hinweg. Es ergibt sich eine durchschnittliche Intercoder-Übereinstimmung von 67,70 Prozent (Range: 56,76 – 75,65).

Tabelle 5.8 Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu **allen Codes**

Thema	Video	Anzahl der Codierungen		prozentuale Übereinstimmung	Cohens Kappa	Konsensuelles Codieren (Anzahl)
		Person 1	Person 2			
NW	D	452	439	72,05	0,71	489
	A	439	384	56,76	0,56	452
TA	F	345	326	61,10	0,60	372
	C	348	333	61,97	0,61	378
TL	B	339	360	73,82	0,73	380
	E	346	368	72,55	0,72	400
	G	445	475	75,65	0,75	494
Gesamt				67,70	0,67	

Der zufallsbereinigte Koeffizient Cohens Kappa beläuft sich auf durchschnittlich $\kappa = 0,67$ (Range: $0,56 < \kappa < 0,75$).¹²⁴ Nach Kuckartz (2016, S. 210) können diese mit

¹²⁴ Kuckartz (2016, S. 216) stellt bei der qualitativen Inhaltsanalyse die Verwendung eines zufallsbereinigten Koeffizienten in Frage, sofern vorab keine Codiersegmente festgelegt wurden. In der vorliegenden Studie wurde zwar auf die Festlegung von Codiereinheiten verzichtet, allerdings liegen

Ausnahme für Video A als gut eingestuft werden. Ein Grund für den niedrigen Wert könnte im vergleichsweise großen Unterschied zwischen der Anzahl der Codierungen beider Personen liegen (Person 1: 439; Person 2: 384).

Da zur Beantwortung verschiedener Forschungsfragen unterschiedliche Kategorien herangezogen werden, erfolgt zusätzlich eine gesonderte Darstellung dieser Untergruppen. Aus Tabelle 5.9 gehen die Intercoder-Übereinstimmungen (ebenfalls für eine Codeüberlappung an Segmentgrenzen von mindestens 70 %) für die Codierungen zu Überzeugungen hinsichtlich des Einsatzes musikbezogener Repräsentationsformen hervor. Die Übereinstimmungswerte liegen im Schnitt bei 77,71 Prozent (Range: 68,75 – 83,40) und ergeben einen zufallsbereinigten Koeffizienten (Cohens Kappa) von durchschnittlich $\kappa = 0,74$ (Range: $0,63 < \kappa < 0,81$). Beide sind nach Kuckartz (2016, S. 210) als gut einzustufen.

Tabelle 5.9 Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu den Codes des Bereichs **Überzeugungen**

Thema	Video	Anzahl der Codierungen		prozentuale Übereinstimmung	Cohens Kappa	Konsensuelles Codieren (Anzahl)
		Person 1	Person 2			
NW	D	42	40	73,17	0,69	39
	A	49	47	68,75	0,63	47
TA	F	32	26	75,86	0,71	31
	C	72	79	76,82	0,72	77
TL	B	28	32	83,33	0,80	34
	E	76	74	82,67	0,79	75
	G	121	114	83,40	0,81	115
Gesamt				77,71	0,74	

Die folgende Tabelle 5.10 zeigt die Werte für alle Codierungen zu den vorher feststehenden Qualitätsmerkmalen (*Adressatenorientierung, Strukturiertheit, Sprech- und Körperausdruck, Sprachliche Verständlichkeit* und *Einsatz musikbezogener Repräsentationen*). Die durchschnittliche Übereinstimmung zwischen beiden Codierenden liegt bei 66,84 (Range: 56,80 – 75,06). Der durchschnittliche Koeffizient (Cohens Kappa) ergibt $\kappa = 0,66$ (Range: $0,55 < \kappa < 0,74$). Im Vergleich zu den Werten des gesamten Kategoriensystems (siehe Tab. 5.8) fallen die Werte des zufallsbereinigten Koeffizienten für die Codierungen zu den Überzeugungen (siehe Tab. 5.9) deutlich niedriger aus. Dies ist insbesondere auf die geringere Anzahl an Kategorien zurückzuführen. Während auch hier die meisten Werte als gut einzustufen sind (Kuckartz, 2016, S. 210), fallen die Ergebnisse der Videos D (NW_o) und C (TA_a) etwas schlechter aus.

vorgegebene Segmente gerade bei den Antworten im Umfang eines Worts oder weniger Worte automatisch vor. Daher werden auch die zufallsbereinigten Koeffizienten (Cohens Kappa) angegeben.

Tabelle 5.10 Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu den Codes des Bereichs **Qualitätsmerkmale I**

Thema	Video	Anzahl der Codierungen		prozentuale Übereinstimmung	Cohens Kappa	Konsensuelles Codieren (Anzahl)
		Person 1	Person 2			
NW	D	258	250	74,80	0,74	282
	A	222	190	56,80	0,55	229
TA	F	157	192	62,46	0,61	202
	C	164	156	56,88	0,55	181
TL	B	192	197	75,06	0,74	201
	E	138	147	70,18	0,69	155
	G	175	207	71,73	0,71	207
Gesamt				66,84	0,66	

Tabelle 5.11 stellt die Inter-coder-Übereinstimmungswerte zu weiteren induktiv entwickelten Kategorien potenzieller Qualitätsmerkmale dar. Diese Kategorien beziehen sich insbesondere auf inhaltliche Aspekte zum Erklärgegenstand (z. B. die Kategorien *fachliche Korrektheit* oder *Menge an Information*) und wurden in einem weiteren Schritt ebenfalls konsensuell codiert. Die Übereinstimmungswerte beider Codierenden liegen im Schnitt bei 72,04 (Range: 59,57 – 84,85); der zufallsbereinigte durchschnittliche Koeffizient (Cohens Kappa) zeigt einen Wert von $\kappa = 0,68$ (Range: $0,53 < \kappa < 0,82$).

Tabelle 5.11 Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu den Codes des Bereichs **Qualitätsmerkmale II**

Thema	Video	Anzahl der Codierungen		prozentuale Übereinstimmung	Cohens Kappa	Konsensuelles Codieren (Anzahl)
		Person 1	Person 2			
NW	D	34	29	76,19	0,73	31
	A	22	25	59,57	0,53	23
TA	F	26	18	62,22	0,56	25
	C	28	31	74,58	0,71	32
TL	B	33	31	71,88	0,68	35
	E	37	35	75,00	0,72	38
	G	32	34	84,85	0,82	33
Gesamt				72,04	0,68	

Abschließend zeigt Tabelle 5.12 die Inter-coder-Übereinstimmungswerte zu themenbezogenen Kategorien, die entsprechend der drei Erklärthemen (Notenwerte, Tonleiter und Taktarten) entwickelt wurden. Daher werden die Durchschnittswerte auch pro Erklärthema und nicht wie bisher über alle Videos hinweg angegeben. So weisen die Inter-coder-Übereinstimmungen Werte von im Schnitt 71,16 für Notenwerte, 66,37 für Taktarten und 70,76 für Tonleiter auf. Alle Einzelwerte liegen zwischen 60,00 und 78,38 Prozent. Die durchschnittlichen zufallsbereinigten Koeffizienten (Cohens Kappa) ergeben die Werte von $\kappa = 0,64$ für Notenwerte, $\kappa = 0,59$ für Taktarten und $\kappa = 0,63$ für

Tonleitern. Die entsprechenden Einzelwerte über alle Themen und Videos liegen zwischen 0,51 und 0,73. Einschränkungen ergeben sich demnach für die Videos A ($\kappa = 0,54$; NW_o), C ($\kappa = 0,51$; TA_a) und E ($\kappa = 0,54$; TL_o).

Tabelle 5.12 Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu den Codes des Bereichs *themenbezogene Kategorien*

Thema	Video	Anzahl der Codierungen		prozentuale Übereinstimmung	Cohens Kappa	Konsensuelles Codieren (Anzahl)
		Person 1	Person 2			
NW	D	39	35	78,38	0,73	50
	A	66	56	63,93	0,54	73
Gesamt (NW)				71,16	0,64	
TA	F	45	32	72,73	0,66	43
	C	21	19	60,00	0,51	21
Gesamt (TA)				66,37	0,59	
TL	B	35	42	70,13	0,62	53
	E	33	54	64,37	0,54	64
	G	47	61	77,78	0,73	61
Gesamt (TL)				70,76	0,63	

Zusammenfassend erweisen sich die dargestellten Werte als zufriedenstellend bis gut, allerdings liegen auch einzelne negative Ausreißer vor. Vor diesem Hintergrund und zur Vergleichbarkeit aller Kategorien erfolgte das konsensuelle Codieren (siehe Kuckartz, 2016, S. 211–212) über das gesamte Material und unter Einbezug aller Codierungen. Bei Stellen ohne Überschneidung einigten sich beide Codierende auf eine endgültige Codierung; diese ist Grundlage der folgenden Analysen. Dabei wurde darauf Wert gelegt, dass es zu keiner Situation des Überstimmens durch die forschende Person gekommen ist, sondern im Zweifel neue Kategorien entwickelt wurden oder der Abschnitt unter *Keine Qualitätsmerkmale* codiert wurde. Diese Form des konsensuellen Codierens bedarf der Idee nach keiner Übereinstimmungswerte. Zur Orientierung kann aus den vorherigen Tabellen (siehe Tab. 5.8, Tab. 5.9, Tab. 5.10, Tab. 5.11, Tab. 5.12) abgelesen werden, wie sich die Zahl der Codierungen über den Prozess des konsensuellen Codierens hinweg verändert hat. Das daraus hervorgegangene endgültige *Kategoriensystem 2b* ist in *Anhang F4* dargestellt.

Glaubwürdigkeit

Das Kriterium Glaubwürdigkeit bezieht sich auf die (interne) Validität der offenen Antworten (siehe Kuckartz, 2016, S. 204–205). Da dieses insbesondere im Kontext von Datenerhebungen mittels Interviews steht, kann es primär als Einschätzung über die Glaubwürdigkeit der Interviewaussagen der befragten Person verstanden werden. Möglichst ausgeschlossen werden sollen Aspekte wie soziale Erwünschtheit, worunter die Glaubwürdigkeit der Antworten leiden könnte. In der vorliegenden Studie beantworteten die

Studienteilnehmer*innen die offenen Fragen zum einen freiwillig und zum anderen ohne Kontakt mit der forschenden Person. Die anonyme Datenerhebung kann zudem zur Steigerung der Glaubwürdigkeit in den offenen Antworten beitragen.

Einschränkungen im Kontext der Validität

Im Gegensatz dazu können Aspekte wie eine inhaltliche Tiefe der Aussagen (Kuckartz, 2016, S. 204–205) aufgrund der fehlenden Möglichkeit zur Nachfrage nicht gewährleistet werden. Auch können keine Verständnisfragen gestellt werden, weshalb das unter anderem von Mayring (2016, S. 147) empfohlene Kriterium der kommunikativen Validierung aufgrund der Form der Datenerhebung ausscheidet. Der Aspekt der argumentativen Interpretationsabsicherung (Mayring, 2016, S. 145), beispielsweise durch das Aufsuchen von Negativfällen, kann ebenfalls nicht gewährleistet werden. Der Grund dafür ist, dass aus den Antworten nicht immer hervorgeht, ob diese positiv oder negativ zu verstehen sind. Auch das Kriterium Nähe zum Gegenstand (Mayring, 2016, S. 146) ist stark eingeschränkt, da die Erhebung nicht in einer realen Unterrichtssituation, sondern in einer Laborsituation mit kurzen Videostimuli stattfand. All diese Limitationen hinsichtlich der Validität der erhobenen offenen Antworten ergeben sich vor allem aus der Form der Datenerhebung, müssen jedoch hinsichtlich der Einschätzung der Güte bei der Ergebnisdarstellung berücksichtigt werden. Mit Blick auf die externe Studiengüte sind nach Kuckartz (2016, S. 203) die Kriterien der inneren Studiengüte ausschlaggebend. Da in der vorliegenden Studie die genannten Limitationen vorliegen, bestehen auch mit Blick auf die Übertragbarkeit der qualitativen Ergebnisse Einschränkungen.

Trotz der genannten Limitationen (siehe auch Kap. 4.9.2) werden die Auswertungen der offenen Antworten in die Ergebnisdarstellung einbezogen. Es wird erwartet, dass die Daten wertvolle Hinweise zu folgenden Aspekten geben können: Als Ergänzung zu den quantitativ analysierten Daten können sie zur Validierung der theoretisch postulierten Qualitätsmerkmale beitragen und grundlegende Überzeugungen hinsichtlich des Einsatzes musikbezogener Repräsentationsformen von Studierenden und Lehrenden aufzeigen. Der explorative Ansatz kann zudem weitere potenzielle Kandidaten für Qualitätsmerkmale guten Erklärens im Musikunterricht hervorbringen und somit auch blinde Flecken beleuchten, die sich aus den theoretischen Setzungen der vorgegebenen Qualitätskriterien ergeben.

5.3 Unterschiede in der Bewertung der Erklärqualität

Dieser Abschnitt berichtet die Ergebnisse zu Unterschieden in der Bewertung der Erklärqualität. Untersucht werden dafür in erster Linie die Globalurteile, die im ersten Teil des Onlinefragebogens von den Studienteilnehmer*innen für jedes Video als Schulnote abgegeben wurden. Fokussiert werden außerdem zwei Faktoren: Zum einen unterscheiden sich die verwendeten Erklärmodi systematisch über alle Videos hinweg, zum anderen liegen Bewertungen von drei unterschiedlichen Statusgruppen vor.

Um aufzuzeigen, welcher Anteil an Unterschieden in der Bewertung auf diese beiden Faktoren zurückzuführen ist, werden für jedes Erklärthema (Notenwerte, Taktarten und Tonleitern) Varianzanalysen mit gemischtem Design (*mixed ANOVA*) als zentrales Analyseverfahren durchgeführt. Die mehrmaligen Bewertungen zu den einzelnen Erklärungen werden als wiederholte Messung verstanden (Messwiederholungsfaktor), die Statusgruppen als dreistufiger Zwischensubjektfaktor. Zudem fungieren das *Globalurteil* als abhängige Variable und die beiden Faktoren, der in den Videos eingesetzte *Erklärmodus* und die Zugehörigkeit der Studienteilnehmer*innen zu einer *Statusgruppe*, als unabhängige Variablen.

Die für die Analyse notwendigen Voraussetzungen innerhalb der entsprechenden Daten werden anhand des *Kolmogorov-Smirnov*-Tests (Test auf Normalverteilung) sowie des *Levene*-Tests (Test auf Varianzhomogenität) überprüft. Bei fehlender Normalverteilung werden nach Bühner und Ziegler (2017, S. 561–562) dennoch parametrische Verfahren angewendet (siehe Kap. 4.9.1). Bei Verletzung der Varianzhomogenität werden dennoch Varianzanalysen durchgeführt, da ausreichend große Stichproben vorliegen – eine mögliche Überschätzung der Effekte wird bei der Interpretation der Ergebnisse berücksichtigt (siehe Bortz & Schuster, 2010, S. 214).

Damit gibt dieser Abschnitt Antworten auf die Forschungsfragen F2, F3 und F4. Die Darstellung der Ergebnisse folgt einem Top-Down-Prinzip. Nach dem Bericht der *mixed ANOVA*, der auch auf mögliche Interaktionseffekte zwischen beiden Faktoren eingeht (Kap. 5.3.1), folgen weiterführende Analysen für die beiden Faktoren Statusgruppe (Kap. 5.3.2) und Erklärmodus (Kap. 5.3.3). Vom Ausgangspunkt des linearen Modells ist es somit einerseits möglich, mittels *Post-hoc*-Analysen spezifische Ergebnisse der beiden Faktoren übersichtlich darzustellen. Andererseits können so auch Ergebnisse aus inhaltsanalytischen Auswertungen der offenen Antworten zielgerichtet einfließen, die insbesondere für den Faktor Erklärmodus zusätzliche Befunde offenlegen können.

5.3.1 Varianzanalyse mit gemischtem Design

F2) Wie unterscheiden sich die Bewertungen von Erklärqualität unter Berücksichtigung von Erklärmodus und Statusgruppe?

Zur Beantwortung dieser Fragen werden zunächst die Unterschiede in den durchschnittlichen Bewertungen pro Video beziehungsweise getrennt nach Erklärthema und zwischen den drei Statusgruppen beschrieben. Anschließend werden inferenzstatistische Verfahren durchgeführt.

Deskriptive Ergebnisse

Zunächst werden die jeweiligen Globalurteile, die im Anschluss als abhängige Variable in die Varianzanalyse einfließen, auf deskriptiver Ebene berichtet. Tabelle 5.13 zeigt die arithmetischen Mittelwerte (*M*) und Standardabweichungen (*SD*) für alle Videos (A bis

G). Zudem wird nach den jeweils verwendeten Erklärmodi (optisch = opt.; akustisch = aku.; optisch und akustisch = opt./aku.) differenziert. Die Gruppierung erfolgt nach Erklärthemen (Notenwerte, Taktarten und Tonleiter). In der oberen Hälfte werden die Werte getrennt nach Statusgruppen, in der unteren Zeile für die gesamte Stichprobe aller Studienteilnehmer*innen berichtet.

Tabelle 5.13 Mittelwerte und Standardabweichungen (getrennt nach Erklärmodus, Thema und Statusgruppe) der Globalurteile zur Erklärqualität

Erklärthema	Notenwerte (NW)		Taktarten (TA)		Tonleiter (TL)		
	Erklärmodus	A: opt. D: opt./aku.	C: aku. F: opt./aku.	E: opt. G: aku. B: opt./aku.			
Videokürzel	NW_o	NW_oa	TA_a	TA_oa	TL_o	TL_a	TL_oa
Statusgruppe	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Schüler*innen	1,90	1,65	2,51	2,32	2,15	2,61	2,10
($N_{max} = 148$)	(1,01)	(0,81)	(1,10)	(1,21)	(1,22)	(1,27)	(1,07)
Studierende	1,82	1,70	2,71	1,57	1,70	2,86	1,45
($N_{max} = 61$)	(0,72)	(0,62)	(1,05)	(0,82)	(0,62)	(1,01)	(0,69)
Lehrende	2,35	2,36	3,06	2,19	2,51	3,67	2,07
($N_{max} = 62$)	(0,97)	(1,00)	(1,21)	(1,07)	(1,02)	(1,22)	(1,02)
Gesamt	1,99	1,83	2,68	2,12	2,13	2,91	1,95
($N_{max} = 271$)	(0,96)	(0,88)	(1,13)	(1,14)	(1,10)	(1,27)	(1,02)

Bem.: opt.: optisch; aku.: akustisch; M: arithm. Mittel; SD: Standardabweichung

Die präsentierten Werte entsprechen einer Schulnotenskala. Üblicherweise werden für Vergleiche statistischer Kennwerte Wörter wie *höherer* beziehungsweise *niedriger* an Stelle von *besser* beziehungsweise *schlechter* verwendet, um keine Wertungen vorzunehmen. Im vorliegenden Fall wäre es jedoch kontraintuitiv, von *niedrigeren Werten* an Stelle von *besseren Noten* zu sprechen, weshalb diese Konvention für die folgenden Ausführungen aufgegeben wird.

Die durchschnittlichen Globalurteile aller Studienteilnehmer*innen pro Video ergeben Werte zwischen 1,83 (NW_oa)¹²⁵ und 2,91 (TL_a). Es fällt auf, dass die Bewertungen der beiden Erklärungen zum Thema Notenwerte (NW_o 1,99; NW_oa: 1,83) tendenziell besser ausfallen als diejenigen zu den Themen Taktarten (TA_a: 2,68; TA_oa: 2,12) und Tonleiter (TL_o: 2,13; TL_a: 2,91; TL_oa: 1,95). Dabei ist zu berücksichtigen, dass in allen drei Varianten der Videoreihenfolge (siehe Kap. 4.3.1) entweder Video A (NW_o) oder Video D (NW_oa) zuerst gezeigt wurden. Da bei der allerersten Bewertung noch kein Referenzwert für die Studienteilnehmer*innen vorlag, könnten die besseren Ergebnisse zum Teil auf eine wohlwollende Beurteilung zu Beginn des Onlinefragebogens zurückzuführen sein. Außerdem werden Notenwerte bereits in der Primarstufe (Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2014) thematisiert, weshalb ein größeres (Vor-)Wissen bei den Schüler*innen zu erwarten ist. Auch kann

¹²⁵ Wie in Kapitel 4.3.1 bereits dargelegt führen die Abkürzungen zunächst das Erklärthema und danach die verwendete(n) Repräsentationsform(en) auf. Das Beispiel NW_oa entspricht Video D mit Erklärthema Notenwerte (NW) und dem kombinierten Erklärmodus (oa) aus optischen (o) und akustischen (a) Repräsentationen. TL_a entspricht demnach Video G; das Erklärthema ist Tonleiter, bezüglich des Erklärmodus werden akustische Repräsentation verwendet – optische Repräsentationen sind in diesem Fall ausgespart.

vermutet werden, dass das Thema im Vergleich zu den anderen beiden die geringste Komplexität aufweist und die Schüler*innen es am leichtesten verstanden haben. Dies kann auch dazu führen, die Erklärung als *besser* – da verständlicher – wahrzunehmen. Darüber hinaus fällt auf, dass die Varianz in den Bewertungen der Studienteilnehmer*innen tendenziell umso geringer ausfällt, je besser die Globalurteile sind – diese Beobachtung wird im weiteren Verlauf noch einmal aufgegriffen.

Unabhängig von den Erklärthemen lassen sich bei den Gesamtglobalurteilen vor allem zwei Aspekte beobachten: Erstens erhalten die Erklärungen mit optischer und akustischer Repräsentationsform bessere Bewertungen (NW_oa: 1,83; TA_oa: 2,12; TL_oa: 1,95) als die Erklärungen, bei denen zum selben Thema nur eine Repräsentationsform verwendet wird (NW_o: 1,99; TA_a: 2,68; TL_o: 2,13; TL_a: 2,91). Zweitens zeigt sich bei den Erklärungen mit nur einer Repräsentationsform, dass die beiden Videos mit ausschließlich akustischer Repräsentationsform (TA_a: 2,91; TL_a: 2,68) schlechter bewertet werden als Erklärungen mit optischer (NW_o: 1,99; TL_o: 2,13). Zudem weisen sie deutlich größere Unterschiede in den Bewertungen im Gegensatz zu den Videos mit kombiniertem Erklärmodus auf (NW_oa: 1,83; TA_oa: 2,12; TL_oa: 1,95).

Diese Tendenzen sind mit wenigen Ausnahmen auch innerhalb der drei Teilstichproben zu beobachten. In allen drei Gruppen (Schüler*innen, Studierende und Lehrende) werden über alle Themen hinweg auch immer die Erklärungen mit optischen und akustischen Repräsentationen am besten bewertet – eine Ausnahme stellen die durchschnittlichen Globalurteile der Lehrenden beim Thema Notenwerte (NW_o: 2,35; NW_oa: 2,36) dar, die sich nur marginal unterscheiden. In allen Gruppen zeigen sich zudem die deutlich schlechteren Bewertungen für die Erklärungen mit ausschließlich akustischer Repräsentation. Eine Ausnahme stellen die Urteile der Schüler*innen für beide Erklärungen zum Thema Taktarten dar: Sie bewerten die Erklärung mit ausschließlich akustischer Repräsentationsform (TA_a: 2,32) zwar ebenfalls schlechter als die Erklärung mit einem kombinierten Erklärmodus (TA_oa: 2,51), allerdings liegen keine so deutlichen Unterschiede in den durchschnittlichen Noten im Vergleich zu den entsprechenden Erklärungen zum Thema Tonleitern (TL_a: 2,61; TL_oa: 2,10) oder zu den anderen Gruppen vor. Auffällig sind zudem die Globalurteile der beiden Erklärungen zum Thema Notenwerte: Im Vergleich zu den anderen beiden Gruppen unterscheiden sich die Bewertungen deutlich voneinander.

Die Varianz der Globalurteile liegt für die Gesamtstichprobe zwischen 0,88 (NW_oa) und 1,27 (TL_a). Auch hier ist analog zu den tendenziell besseren Globalurteilen auffällig, dass die Standardabweichungen bei den beiden Erklärungen zu Notenwerten (NW_o: 0,96; NW_oa: 0,88) niedriger ausfallen als bei den Videos zu Taktarten (TA_a: 1,13; TA_oa: 1,14) und Tonleiter (TL_o: 1,10; TL_a: 1,27; TL_oa: 1,02). Beim Vergleich der drei Statusgruppen fällt auf, dass die Standardabweichungen der Gruppe der Studierenden zwischen 0,62 (NW_oa und TL_o) und 1,05 (TA_a) liegen; sie fallen damit niedriger aus als die Standardabweichungen der Schüler*innen [min = 0,81; max = 1,27] und den Lehrenden [min = 0,97; max = 1,22]. Das bedeutet, dass die Gruppe der Studierenden tendenziell einheitlichere Bewertungen als die anderen Gruppen abgibt und sich womöglich

homogener zusammensetzt. Außerdem sind diese Unterschiede bei der Einordnung der Ergebnisse zu berücksichtigen, da sie zu Verzerrungen führen können.

Neben dieser statusgruppenspezifischen Beobachtung können Zusammenhänge dahingehend vermutet werden, dass mit besseren Globalurteilen tendenziell eine geringere Varianz in der Bewertung einhergeht. Das bedeutet gleichzeitig auch: Je schlechter der Durchschnittswert der Globalbewertung wird, desto mehr Uneinigkeit ist in den Bewertungen der Studienteilnehmer*innen festzustellen. Beim Blick auf alle 21 nach Erklärung und Statusgruppe getrennten Mittelwerte und Standardabweichungen wird deutlich, dass die Höhe der Globalbewertungen sowie die der jeweiligen Standardabweichungen über alle Statusgruppen hinweg signifikant sind und hoch miteinander korrelieren ($r_{\text{Spearman}} = 0,71$; $p < 0,01$). Je höher – und damit *schlechter* – die Globalurteile ausfallen, desto mehr weichen die einzelnen Bewertungen im Schnitt vom arithmetischen Mittel ab. Diese Beobachtung gilt es bei der Interpretation der Ergebnisse zu berücksichtigen.

Varianzanalysen mit gemischtem Design

Im Folgenden werden die Ergebnisse der Varianzanalysen mit gemischtem Design berichtet. Die Prüfung der Voraussetzungen ergab an mehreren Stellen Verletzungen: Einige Variablen (gemeint sind hier Bewertungen pro Video und nach Statusgruppen getrennt) weichen signifikant von einer Normalverteilung ab (alle $p < 0,05$; siehe *Anhang D1*). Aufgrund der Stichprobengröße werden dennoch robuste Ergebnisse erwartet (siehe Bortz & Schuster, 2010, S. 214). Zudem liegen für die Videos NW_o, NW_oa, TA_oa, TL_o und TL_oa keine homogenen Varianzen vor (alle $p < 0,05$; siehe *Anhang D2*). Dadurch ist die Interpretation der Ergebnisse limitiert. Signifikante Abweichungen von der Sphärizität, die für die *mixed ANOVA* der drei Erklärungen zum Thema Tonleiter vorliegen, sind nach *Greenhouse-Geisser* korrigiert (siehe *Anhang D3*). Tabelle 5.14 zeigt die Ergebnisse. Die Darstellung erfolgt für die drei mehrfaktoriellen Varianzanalysen für unabhängige Stichproben entlang der drei Themen Notenwerte, Taktarten und Tonleiter.

Tabelle 5.14 Mittelwerte (Standardabweichungen) der Globalurteile pro Video und Statusgruppe; je Erklärthema eine Varianzanalyse mit gemischtem Design sowie den Faktoren Erklärmodus und Statusgruppe

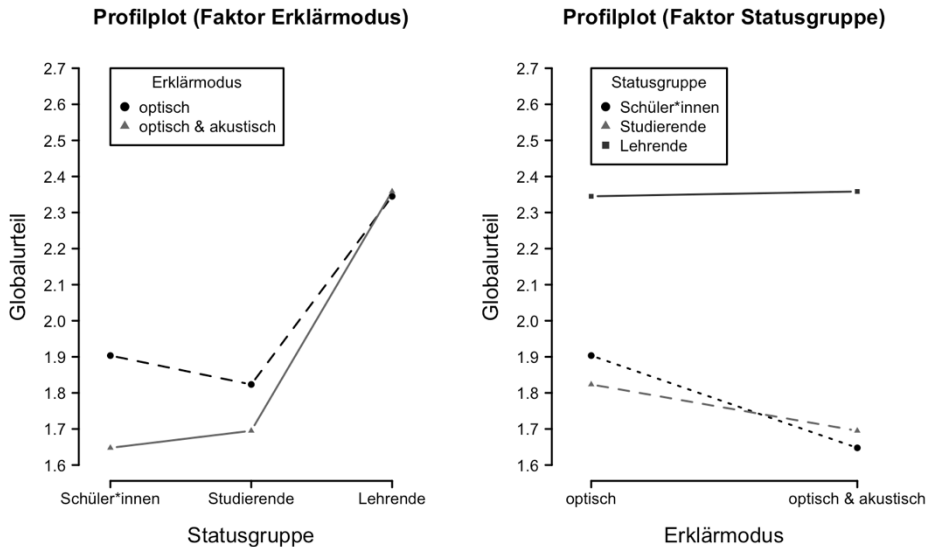
Erklärthema	Notenwerte			Taktarten			Tonleiter		
	Erklärmodus	A: opt.	D: opt./aku.	C: aku.	F: opt./aku.	E: opt.	G: aku.	B: o./a.	
Videokürzel	NW_o	NW_oa	TA_a	TA_oa	TL_o	TL_a	TL_oa		
Statusgruppe	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	
Schüler*innen ($N_{max} = 148$)	1,90 (1,01)	1,65 (0,81)	2,51 (1,10)	2,32 (1,21)	2,15 (1,22)	2,61 (1,27)	2,10 (1,07)		
Studierende ($N_{max} = 61$)	1,82 (0,72)	1,70 (0,63)	2,71 (1,05)	1,56 (0,83)	1,70 (0,63)	2,86 (1,02)	1,45 (0,69)		
Lehrende ($N_{max} = 62$)	2,35 (0,97)	2,36 (0,97)	3,09 (1,19)	2,21 (1,06)	2,50 (1,03)	3,64 (1,23)	2,09 (1,02)		
ANOVA	<i>df</i>	<i>F</i>	η^2	<i>df</i>	<i>F</i>	η^2	<i>df</i>	<i>F</i>	η^2
Erklärmodus	1	2,85	0,01	1	81,07**	0,24	2	108,22**	0,29
Statusgruppe	2	16,45**	0,11	2	4,48*	0,03	2	12,01**	0,08
Mod. x Statusg.	2	1,33	0,01	2	23,28**	0,11	4	10,83**	0,08
Residuen	262			259			259/518		

Bem.: N_{max} : aufgrund fehlender Werte bei einzelnen Videos den Berechnungen pro Statusgruppe maximal zugrunde liegende Personenanzahl; *M*: arithm. Mittel; *SD*: Standardabweichung; ANOVA: Varianzanalyse mit gemischtem Design; *df*: Freiheitsgrade; *F*: *F*-Wert; η^2 : partielles Eta-Quadrat; *: $p \leq 0,05$; **: $p \leq 0,01$.

Notenwerte

Die Varianzanalyse der Erklärungen zum Thema Notenwerte (siehe Tab. 5.14) ergibt, dass keine signifikanten Unterschiede in der durchschnittlichen Bewertung der Erklärqualität auf die verschiedenen Erklärmodi zurückgeführt werden können ($F(1, 262) = 2,85$; n. S. ; $\eta^2 = 0,01$). Es wird vermutet, dass die zusätzliche akustische Repräsentation in Video NW_oa – es wurden unterschiedlich lange Notenwerte am Klavier vorgespielt – zu unbedeutend und zu wenig aussagekräftig ist, um im Modell einen signifikanten Effekt hervorzurufen. Demgegenüber liegen signifikante Unterschiede in den durchschnittlichen Globalurteilen zwischen den drei Statusgruppen vor ($F(2, 262) = 16,45$; $p < 0,01$; $\eta^2 = 0,11$). Nach Cohen (1988, Kap. 8) kann dabei ein mittlerer bis großer Effekt hinsichtlich der unterschiedlichen Bewertung von Erklärqualität durch die Zugehörigkeit zu den Gruppen beobachtet werden. Zwischen beiden Haupteffekten liegen im Modell keine signifikanten Interaktionen vor ($F(2, 262) = 1,33$; n. S. ; $\eta^2 = 0,01$). Für den Faktor Statusgruppe werden in weiterführenden Analysen mittels *Post-hoc*-Tests die genauen Unterschiede zwischen den drei Gruppen überprüft (siehe Kap. 5.3.2).

Abbildung 5.2 Profilplots zu den Globalurteilen der Faktoren Erklärmodus und Statusgruppe (Thema Notenwerte)



Der Blick auf die beiden Profilplots (siehe Abb. 5.2) zeigt, dass der Einsatz einer akustischen Repräsentation im Video NW_oa gegenüber der Erklärung in NW_o, die darauf verzichtet, im statistischen Modell zu keinem signifikanten Einfluss auf das Globalurteil führt. Gleichwohl lassen sich deskriptiv bei der Gruppe der Schüler*innen größere Unterschiede in den Bewertungen hinsichtlich der verschiedenen Erklärmodi erkennen.

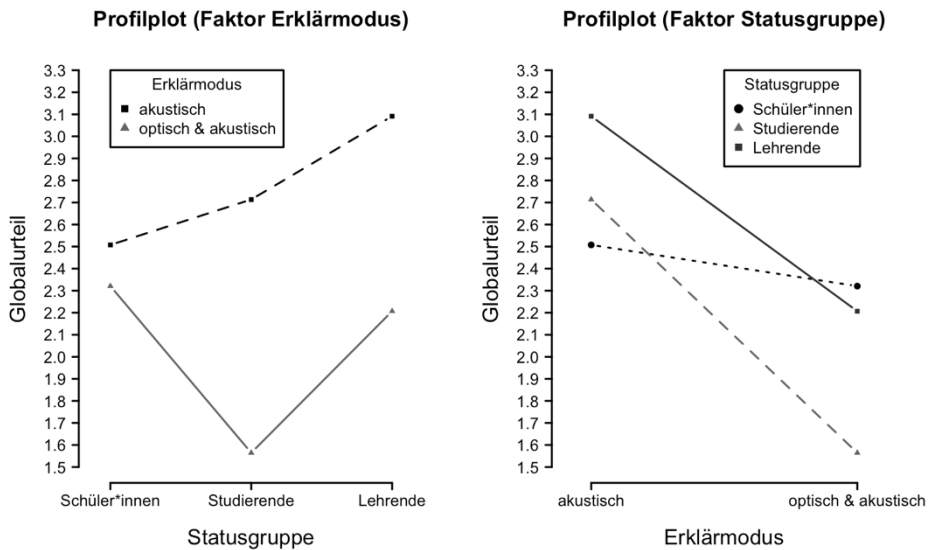
Taktarten

Die Varianzanalyse der beiden Erklärungen zum Thema Taktarten (siehe Tab. 5.14) zeigt einen signifikanten Unterschied in den durchschnittlichen Globalurteilen, der auf den gewählten Erklärmodus zurückzuführen ist ($F(1, 259) = 81,07; p < 0,01; \eta^2 = 0,24$). Mit einem nach Cohen (1988, Kap. 8) starken Effekt ($\eta^2 = 0,24$) kann nahezu ein Viertel der Varianz in den Globalurteilen dadurch erklärt werden, dass bei der Erklärung in Video TA_a die optische Repräsentation fehlt, während sie Bestandteil der Erklärung TA_oa ist. Auch zeigen sich signifikante Unterschiede zwischen den durchschnittlichen Bewertungen der drei Statusgruppen ($F(2, 259) = 4,48; p < 0,05; \eta^2 = 0,03$). Allerdings ist die sehr kleine partielle Effektstärke ($\eta^2 = 0,03$) ein Indiz dafür, dass die Zugehörigkeit zur Statusgruppe einen eher unbedeutenden Einfluss auf die durchschnittlichen Globalbewertungen nimmt.

Darüber hinaus ergibt die Analyse einen signifikanten, mittelstarken partiellen Interaktionseffekt zwischen beiden Faktoren ($F(2, 259) = 23,28; p < 0,01; \eta^2 = 0,11$). Der Blick auf die Profilanalysen der beiden alternativen Darstellungsvarianten verdeutlicht, dass hybride Interaktionen vorliegen (siehe Abb. 5.3). Beim Faktor Erklärmodus kann davon ausgegangen werden, dass die durchschnittlichen Bewertungen bei Erklärung

TA_a (Erklärmodus: akustisch) unabhängig der Statusgruppenzugehörigkeit höhere Werte als bei Erklärung TA_oa (Erklärmodus: optisch & akustisch) einnehmen und damit hinsichtlich der Erklärqualität schlechter ausfallen. Keine generellen Aussagen sind für den Faktor Statusgruppe möglich, da hier nur bedingte Effekte vorliegen. So wirken die verschiedenen Erklärmodi je nach Zugehörigkeit zur Statusgruppe unterschiedlich stark auf die Erklärqualität: Die Hinzunahme einer optischen Repräsentation beeinflusst das durchschnittliche Urteil zur Erklärqualität der Schüler*innen deutlich geringer als das der Studierenden und der Lehrenden.

Abbildung 5.3 Profilplots zu den Globalurteilen der Faktoren Erklärmodus und Statusgruppe (Thema Taktarten)



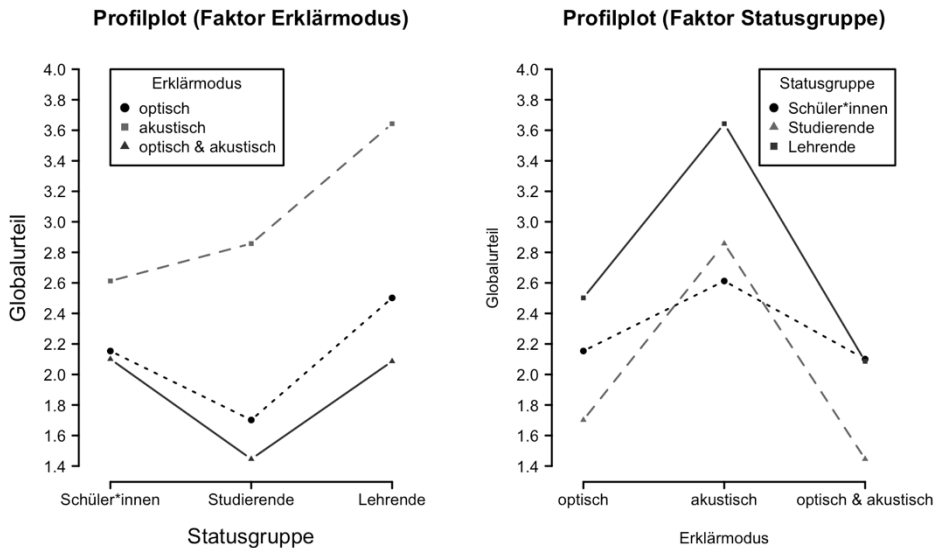
Weiterführende Analysen fokussieren die Interpretation des Faktors Erklärmodus (siehe Kap. 5.3.3), da sich über alle Gruppen zwar unterschiedlich starke, jedoch grundsätzlich konvergierende Vorteile für die Kombination aus optischer und akustischer Repräsentationsform gegenüber des rein akustischen Erklärmodus zeigen.

Tonleiter

Hinsichtlich der drei Erklärungen zum Thema Tonleiter können signifikante Unterschiede in den durchschnittlichen Globalurteilen auf den Faktor Erklärmodus zurückgeführt werden ($F(2, 259) = 108,22; p < 0,01$; siehe Tab. 5.14) und zeigen einen nach Cohen (1988, Kap. 8) sehr starken Effekt ($\eta^2 = 0,29$). Auch kann die Zugehörigkeit zur Statusgruppe einen Teil der die Varianz mit einem mittleren Effekt erklären ($F(2, 259) = 12,01; p < 0,01; \eta^2 = 0,08$), der jedoch gegenüber dem Erklärmodus deutlich geringer ausfällt. Zwischen beiden Faktoren liegt ein signifikanter mittelstarker Interaktionseffekt vor ($F(4, 518) = 10,83; p < 0,01; \eta^2 = 0,08$).

Ähnlich den Erklärungen zu Taktarten weisen die Profilanalysen auch bei denen zu Tonleitern auf hybride Interaktionen hin (siehe Abb. 5.4). Es zeigt sich, dass unabhängig des Faktors Statusgruppe im Schnitt jeweils die besten Bewertungen für den Erklärmodus *optisch* und *akustisch* vorliegen, wobei *optisch* in geringem Maße und *akustisch* deutlich schlechter beurteilt wird. Das deckt sich in etwa mit den Beobachtungen aus den bisherigen Erklärungen zu Notenwerten und Taktarten. Zu prüfen ist in weiterführenden Analysen (siehe Kap. 5.3.3), welche der drei Erklärmodi sich signifikant voneinander unterscheiden. Bei dem Faktor Statusgruppe liegen nur bedingte Effekte vor: Die Schüler*innen bewerten den Erklärmodus *akustisch* besser als die anderen beiden Gruppen und dafür *optisch* und *akustisch* schlechter. So liegen keine einheitlichen Bewertungen der drei Statusgruppen vor, was bei weiterführenden Analysen (siehe Kap. 5.3.2) zu berücksichtigen ist.

Abbildung 5.4 Profilplots zu den Globalurteilen der Faktoren Erklärmodus und Statusgruppe (Thema Tonleiter)



Zusammenfassung

Zusammenfassend ergeben sich aus den drei Varianzanalysen zur Forschungsfrage zwei (F2), wie sich die Bewertungen zur Erklärqualität unter Berücksichtigung der Faktoren Statusgruppe und Erklärmodus unterscheiden, folgende Ergebnisse: Beide Faktoren können Varianz in den Globalurteilen erklären, wobei der Faktor Erklärmodus bei den Erklärungen zu Taktarten und Tonleitern als Haupteffekt zu betrachten ist, wohingegen beim Thema Notenwerte keine signifikanten Unterschiede zu erkennen sind. Der Faktor Statusgruppe kann nur beim Thema Notenwerte als Haupteffekt betrachtet werden; bei den Themen Taktarten und Tonleiter liegen nur bedingte, vom Faktor Erklärmodus abhängige

Effekte vor. Damit wird aufgrund der Größe der Effekte angenommen, dass der Faktor Erklärmodus bedeutsamer für die Bewertungen der Erklärqualität ist als der Faktor Statusgruppe.

In weiterführenden Analysen sollen *Post-hoc*-Tests die Fragen beantworten, welche Erklärmodi sich unabhängig von den Themen signifikant unterscheiden (siehe Kap. 5.3.3). Hinsichtlich der Frage, welche Unterschiede genau zwischen den einzelnen Bewertungen der Statusgruppen im Schnitt vorliegen, werden in den nun folgenden Analysen (siehe Kap. 5.3.2) bei den Erklärungen zu Taktarten und Tonleitern die jeweils konstatierten Interaktionseffekte berücksichtigt.

5.3.2 Faktor Statusgruppe

F3) Wie unterscheiden sich die Bewertungen der Erklärqualität zwischen Schüler*innen, Studierenden und Lehrenden?

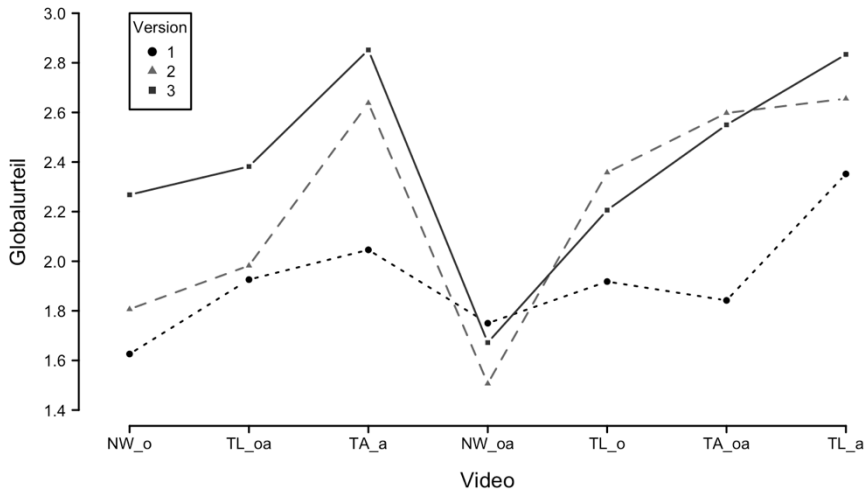
Zu dieser Frage wurden in Kapitel 3.2.2 zwei Hypothesen formuliert. Zum einen wird erwartet, dass zwischen den beiden Perspektiven auf unterrichtliches Erklären – Schüler*innen auf der einen und die Gruppen der Studierenden und Lehrenden auf der anderen Seite – signifikante Unterschiede in den Bewertungen vorliegen (*Hypothese 3.1*). Zum anderen wird angenommen, dass aufgrund unterschiedlicher Expertise auch zwischen Studierenden und Lehrenden signifikante Unterschiede in der Bewertung der Erklärqualität festgestellt werden (*Hypothese 3.2*). Für beide Hypothesen ist zudem von Bedeutung, ob grundsätzlich verschiedene statusgruppenspezifische Bewertungsmuster festzustellen sind – etwa ob eine Statusgruppe grundsätzlich besser bewertet als eine andere und dabei zwar die Bereiche auf der Notenskala differieren, jedoch tendenziell Ähnlichkeiten hinsichtlich der unterschiedlichen Erklärmodi vorliegen. Daher werden zunächst alle drei Statusgruppen für sich betrachtet und kurze Bewertungsprofile erstellt. In diesen werden die Ergebnisse vor dem Hintergrund möglicher Kontrollvariablen und statusgruppenbezogener Versions- oder Reihenfolgeeffekte eingeordnet sowie die deskriptiv besten und schlechtesten Videos vorgestellt. Erst im Anschluss erfolgt ein Vergleich, in dem die Gemeinsamkeiten und Unterschiede beleuchtet und die aufgestellten Hypothesen überprüft werden. Als *Post-hoc*-Analysen folgen abschließend paarweise *t*-Tests mit Alphafehlerkorrektur (*Bonferroni*), um sichtbar zu machen, welche Gruppen genau sich pro Erklärung unterscheiden.

Schüler*innen

Die Unterschiede in den Globalurteilen der Schüler*innen können zum Teil auf die verschiedenen Fragebogenversionen und damit einhergehende Reihenfolgeeffekte zurückgeführt werden. Version 1 wird im Vergleich zu 2 und 3 tendenziell besser bewertet ($F(2, 142) = 5,54; p < 0,01; \eta^2 = 0,07$); damit hat die Fragebogenversion einen mittelstarken Effekt auf die durchschnittliche Bewertung der Globalurteile (siehe Abb. 5.5). Der

signifikante, jedoch sehr schwache Interaktionseffekt zwischen den Versionen und den Videos als Messwiederholungsfaktor kann vernachlässigt werden ($F(12, 852) = 2,60; p < 0,01; \eta^2 = 0,01$).

Abbildung 5.5 Durchschnittliche Globalurteile der Schüler*innen getrennt nach den drei Versionen



Die Bewertungen der Erklärqualität durch die Schüler*innen (siehe Tab. 5.14) erstrecken sich über eine Spannweite von 1,65 (Minimumwert, NW_oa) bis 2,61 (Maximumwert, TL_a). Die zwei am besten bewerteten Videos sind die beiden Erklärungen zu Notenwerten (NW_oa: 1,65; NW_o: 1,90), wobei die Interaktion zwischen Fragebogenversion und Globalurteil bei Video NW_oa die Aussagekraft dieses deskriptiven Befundes limitiert.

Interessant sind in diesem Zusammenhang die Ergebnisse von zwei Kontrollvariablen: Zum einen schätzten die Schüler*innen auf einer sechsstufigen Skala nach jedem Video ein, inwieweit sie das Erklärte verstanden haben; zum anderen gaben sie auf einer dreistufigen Skala an, wie hoch ihr Vorwissen zu den erklärten Inhalten im Video war (siehe Tabelle 5.15). Es wird vermutet, dass die Komplexität des Themas einen Einfluss auf die Globalbewertung der Schüler*innen hat.

Tabelle 5.15 zeigt deskriptiv, dass die Schüler*innen ihren eigenen Angaben nach das größte Vorwissen zum Thema Notenwerte zu besitzen (NW_o: 1,42; NW_oa: 1,33). Das erscheint plausibel, da Notenwerte im Gegensatz zu den anderen beiden Themen bereits in der Primarstufe zum Inhalt des Lehrplans gehören (z. B. Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2014). Vor diesem Hintergrund liegt auch nahe, dass die Schüler*innen ihren Angaben nach die Erklärungen zu Notenwerten besser verstanden haben als zu den anderen Themen (NW_o: 1,90; NW_oa: 1,72). Nach den Konventionen von Cohen (1988, Kap. 3) korreliert die Einschätzung zum vermeintlichen Verstehen der Inhalte hoch mit der Globalbewertung zur Erklärqualität

($t(1013) = 51,99$; $p < 0,01$; $r = 0,85$). Für die Variable *Vorwissen* kann aufgrund vieler fehlender Daten keine Korrelation ausgewiesen werden.

Tabelle 5.15 Kontrollvariablen mit Einschätzungen der **Schüler*innen** zu ihrem themenbezogenen Vorwissen und videobelegten Verstehen

Schüler*innen	Erklärthemen	Notenwerte		Taktarten		Tonleiter		
		NW_o	NW_oa	TA_a	TA_oa	TL_o	TL_a	TL_oa
Vorwissen								
Was im Video erklärt wurde,								
(1) ... wusste ich vorher schon.								
		1,42	1,33	1,88	1,88	1,78	1,94	1,90
(2) ... wusste ich teilweise schon.								
(3) ... wusste ich vorher noch nicht.								
Verstehen								
Was im Video erklärt wurde, habe ich								
(1) ... sehr gut verstanden.								
		1,90	1,72	2,61	2,50	2,30	2,85	2,26
(2) ... gut verstanden.								
(3) ... einigermaßen verstanden.								
(4) ... nicht so gut verstanden.								
(5) ... schlecht verstanden.								
(6) ... sehr schlecht verstanden.								

Bem.: Werte geben jeweils das arithm. Mittel an.

Studierende

Bei der Gruppe der Studierenden unterscheiden sich die Bewertungen unter Berücksichtigung der verschiedenen Fragebogenversionen nicht signifikant voneinander ($F(2, 56) = 0,62$; n. s.). Allerdings liegt ein signifikanter, starker Interaktionseffekt zwischen Version und Videoreihenfolge vor, der einen Teil der Varianz des Globalurteils aufklärt ($F(12, 336) = 4,41$; $p < 0,01$; $\eta^2 = 0,14$). Dazu zeigt Abbildung 5.6 die durchschnittlichen Globalurteile der Studierenden für jede der drei Versionen. Die beiden besten Noten vergeben die Studierenden (siehe Tab 5.15) für zwei Erklärungen zu den Themen Tonleiter und Taktarten mit optischer und akustischer Repräsentation (TL_oa: 1,45; TA_oa: 1,56). Außerdem bewerten sie die beiden Erklärungen, in denen keine optische Repräsentation vorkommt (TL_a: 2,86; TA_a: 2,71), am schlechtesten. Daraus ergibt sich eine Spannweite von 1,45 (Minimumwert, Video TL_oa) und 2,86 (Maximumwert, Video TL_a), die gegenüber den Urteilen der Schüler*innen größer ausfällt.

Abbildung 5.6 Durchschnittliche Globalurteile der Studierenden getrennt nach den drei Versionen

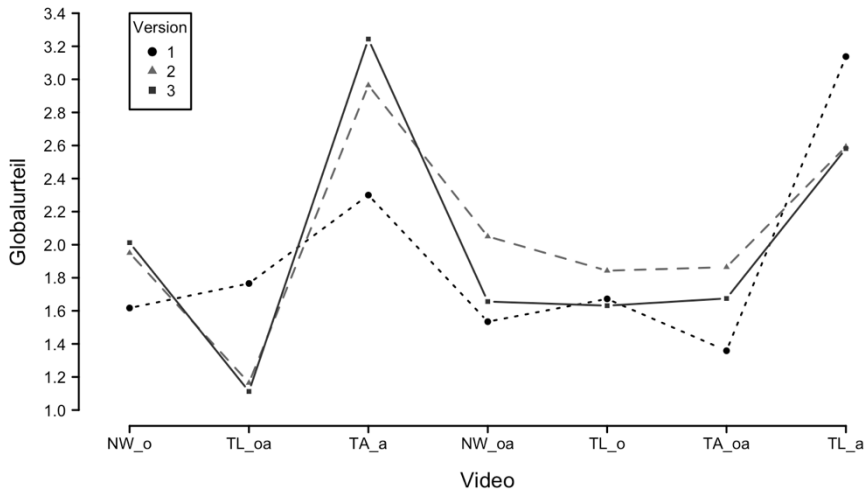


Tabelle 5.16 zeigt, dass auch bei den Studierenden die Bewertungen zur Erklärqualität hoch mit deren Einschätzungen darüber korrelieren, ob Schüler*innen einer sechsten Jahrgangsstufe die Erklärung verstehen können ($t(411) = 32,92; p < 0,01; r = 0,85$). Dabei orientieren sich die Beurteilungen der Studierenden bezüglich der Frage, was Schüler*innen am besten verstehen, weniger an Erklärthemen – im Gegensatz zur Selbsteinschätzung der Schüler*innen –, sondern mehr an einzelnen Erklärmodi: Ein höheres Verstehenspotenzial wird analog der besseren Bewertung der Erklärqualität bei den Erklärungen mit optischer (NW_o, TL_o) sowie optischer und akustischer Repräsentationsform (NW_oa, TA_oa, TL_oa) vermutet. Für die statusgruppenspezifische Bewertung der Studierenden sei an dieser Stelle noch einmal (siehe Kap 5.3.1) auf die auffällige Beobachtung hingewiesen, dass sie im Gegensatz zu Schüler*innen und Lehrenden homogenere Globalurteile vergeben.

Tabelle 5.16 Einschätzungen zum Verstehenspotenzial der Erklärung für Schüler*innen durch **Studierende**

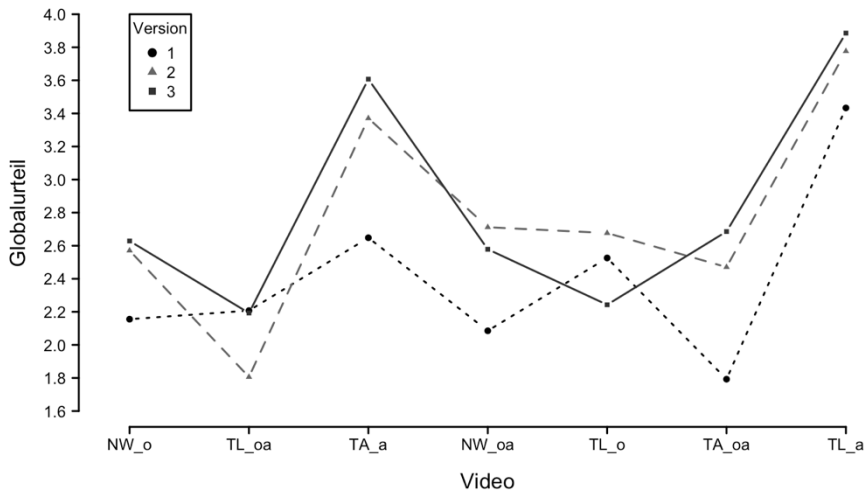
Erklärthemen	Notenwerte		Taktarten		Tonleiter		
	NW_o	NW_oa	TA_a	TA_oa	TL_o	TL_a	TL_oa
Verstehenspotenzial							
Was im Video erklärt wurde, verstehen Schüler*innen einer 6. Jgst.							
(1) ... sehr gut.							
(2) ... gut.							
(3) ... einigermaßen.	2,07	1,79	3,17	1,78	1,68	3,48	1,75
(4) ... nicht so gut.							
(5) ... schlecht.							
(6) ... sehr schlecht.							

Bem.: Werte geben jeweils das arithm. Mittel an.

Lehrende

Auch bei den Lehrenden zeigen sich keine signifikanten Effekte hinsichtlich der verschiedenen Fragebogenversionen ($F(2, 55) = 2,39$; n. s.), wobei signifikante Interaktionseffekte zwischen Version und Videoreihenfolge vorliegen, die einen mittelstarken Einfluss auf die durchschnittliche Bewertung des Globalurteils haben ($F(12, 330) = 1,92$; $p < 0,05$; $\eta^2 = 0,07$). Abbildung 5.7 zeigt die die durchschnittlichen Globalurteile der Lehrenden für alle drei Versionen. Hier kann im Gegensatz zu den Studierenden vermutet werden, dass für die Gruppe der Lehrenden bereits aus dem Berufsalltag mehr Vergleichsmaßstäbe vorliegen, sodass es unerheblich ist, welche Erklärung zuerst bewertet werden soll.

Abbildung 5.7 Durchschnittliche Globalurteile der Lehrenden getrennt nach den drei Versionen



Die Gruppe der Lehrenden bewertet die Erklärungen am besten (siehe Tab. 5.14), die bei den Themen Tonleiter und Taktarten eine optische und akustische Repräsentation einsetzen (TL_oa: 2,09; TA_oa: 2,21). Die geringste Erklärqualität wird den beiden Erklärungen attestiert, die zwar akustische, jedoch keine optische Repräsentation verwenden (TL_a: 3,64; TA_a: 3,09). Die Spannweite erstreckt sich dadurch von 2,09 (Minimumwert, Video TL_oa) bis 3,64 (Maximumwert, Video TL_a).

Auch spielen bei den Lehrenden die Einschätzungen des Verstehenspotenzials der Erklärung für die Schüler*innen eine wichtige Rolle und korrelieren hoch mit der Bewertung der Erklärqualität ($t(404) = 27,65$; $p < 0,01$; $r = 0,81$). Im Gegensatz zur Selbsteinschätzung der Schüler*innen, die vor allem bei den Erklärungen zu Notenwerten angeben, am meisten verstanden zu haben, vermuten neben den Studierenden auch die Lehrenden, dass insbesondere die Erklärungen mit optischer (NW_o, TL_o) sowie optischer und akustischer Repräsentationsform (NW_oa, TA_oa, TL_oa) mehr Potenzial zum Verstehen bereithalten (siehe Tab. 5.17). Während die Studierenden bei diesen Videos jedoch

auch davon ausgehen, dass Schüler*innen einzelne Erklärungen gut bis sehr gut verstehen können, fällt die Prognose der Lehrenden schlechter aus: Die Videos TA_oa und TL_oa werden zwischen *einigermaßen zu verstehen* und *gut zu verstehen* eingeordnet.

Tabelle 5.17 Einschätzungen zum Verstehenspotenzial der Erklärung für Schüler*innen durch **Lehrende**

Lehrende	Erklärthemen	Notenwerte		Taktarten		Tonleiter		
		NW_o	NW_oa	TA_a	TA_oa	TL_o	TL_a	TL_oa
Verstehenspotenzial								
Was im Video erklärt wurde, verstehen Schüler*innen einer 6. Jgst.								
(1) ... sehr gut.								
(2) ... gut.								
		2,37	2,42	3,39	2,31	2,58	4,12	2,31
(3) ... einigermaßen.								
(4) ... nicht so gut.								
(5) ... schlecht.								
(6) ... sehr schlecht.								

Bem.: Werte geben jeweils das arithm. Mittel an.

Vergleich der drei Statusgruppen

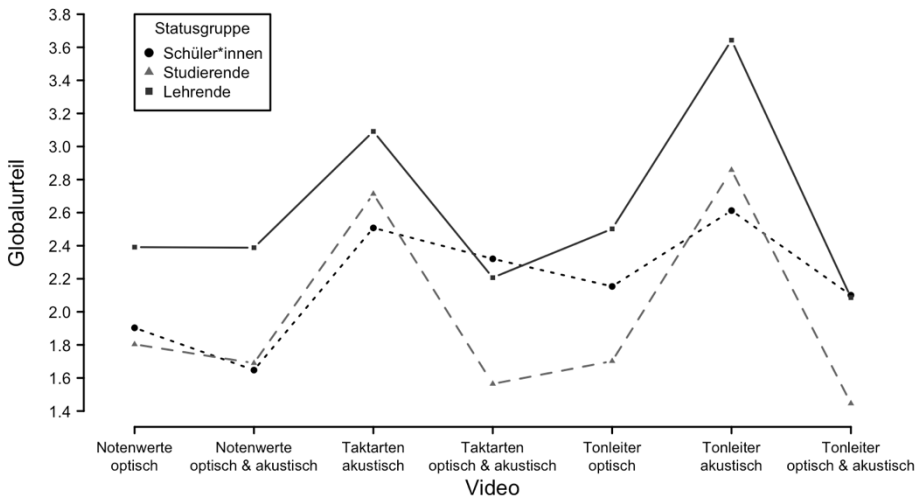
Um die Frage zu beantworten, wie sich die Globalurteile zur Erklärqualität zwischen Schüler*innen, Studierenden und Lehrenden unterscheiden, werden die bisher für alle drei Statusgruppen getrennt dargestellten Befunde zusammengeführt und miteinander verglichen.

Gemein ist allen drei Gruppen, dass sowohl die Angaben der Schüler*innen zu ihrem Verstehen als auch die Einschätzungen der Studierenden und Lehrenden, wie verständlich das im Video Erklärte war, jeweils hoch mit den Bewertungen der Erklärqualität korrelieren. Damit erhält die Zielperspektive *Verstehen* eine bedeutende Rolle für die Qualität von Erklärungen. Inwieweit für die Schüler*innen Zusammenhänge mit ihrem eigenen Vorwissen vorliegen, kann mit Hilfe der erhobenen Daten nur bedingt beantwortet werden. Zwar lässt sich deskriptiv anhand der Selbsteinschätzungen der Schüler*innen zeigen, dass das Vorwissen zu Notenwerten höher ausfällt. Dies führt jedoch nicht zu einem signifikanten Einfluss hinsichtlich der Bewertungen der Globalurteile. Auch können anhand der erhobenen Daten keine generellen Aussagen über die Wirksamkeit der Erklärungen getroffen werden. Erste Hinweise diesbezüglich sollen jedoch die Ergebnisse eines kurzen Vorwissenstest liefern (siehe Kap. 5.5). Alle drei Gruppen sind sich darin einig, dass die Erklärungen ohne optische Repräsentation die geringste Qualität aufweisen.

Unterschiede zeigen sich zwischen den drei Gruppen sowohl hinsichtlich der Spannweite zwischen der jeweiligen besten und schlechtesten Bewertung als auch bezüglich des gewählten Abschnitts auf der Notenskala. Während Aussagen darüber, wer *bessere* oder *schlechtere* Noten vergibt, aufgrund der gewählten Perspektiventriangulation weder sinnvoll noch aussagekräftig sind, lässt der unterschiedliche Range vermuten, dass insbesondere die Lehrenden (Range = [2,09;3,64], 1,55) und mit geringerem Abstand auch die

Studierenden (Range = [1,45;2,85], 1,40) gegenüber den Schüler*innen (Range = [1,65;2,61], 0,96) etwas differenziertere Bewertungen vornehmen. Studierende und Lehrende bewerten die Qualität der einzelnen Erklärungen in absoluten Globalurteilen zwar unterschiedlich, allerdings lassen sich ähnliche Bewertungstendenzen erkennen. Die Linien verlaufen jeweils mit Abstand, jedoch in der Tendenz parallel zueinander. Bei den Bewertungen der Schüler*innen sind insgesamt nicht nur weniger Ausschläge zu beobachten, sie unterscheiden sich beispielsweise auch bei den Erklärungen zu Taktarten deutlich von denen der anderen beiden Gruppen: Schüler*innen sehen im Gegensatz zu Studierenden und Lehrenden keine großen Unterschiede in der Erklärqualität (siehe Abb. 5.8). Vor dem Hintergrund dieser Beobachtungen werden abschließend die Ergebnisse der zugehörigen *Post-hoc*-Tests vorgestellt. Sie dienen dazu, *Hypothese 3.1* und *Hypothese 3.2* zu überprüfen.

Abbildung 5.8 Globalurteile pro Video getrennt nach Statusgruppen



Post-hoc-Tests

Tabelle 5.18 führt nochmals die Mittelwerte und Standardabweichungen zu den einzelnen Videos getrennt nach den drei Statusgruppen auf und berichtet zudem die Ergebnisse der paarweisen Vergleiche. Alle *Post-hoc*-Tests wurden mittels *Bonferroni*-Korrektur auf ein Alphaniveau von $p = 0,05$ korrigiert. Bei den Videos NW_o, NW_oa und TL_o wurden aufgrund der Verletzung der Varianzhomogenität *Welch*-Tests anstelle der klassischen *t*-Tests durchgeführt.

Tabelle 5.18 Mittelwerte und Standardabweichungen zu allen Erklärungen getrennt nach Statusgruppen sowie Ergebnisse der *Bonferroni*-korrigierten *Post-hoc*-Analysen

Erklärthema Erklärmodus Videokürzel	Notenwerte		Taktarten		Tonleiter		
	opt. NW_o	opt. & aku. NW_oa	aku. TA_a	opt. & aku. TA_oa	opt. TL_o	aku. TL_a	opt. & aku. TL_oa
Statusgruppe	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Schüler*innen (N _{max} = 148)	1,90 (1,01)	1,65 (0,81)	2,51 (1,10)	2,32 (1,21)	2,15 (1,22)	2,61 (1,27)	2,10 (1,07)
Studierende (N _{max} = 61)	1,82 (0,72)	1,70 (0,63)	2,71 (1,05)	1,56 (0,83)	1,70 (0,63)	2,86 (1,02)	1,45 (0,69)
Lehrende (N _{max} = 62)	2,35 (0,97)	2,36 (0,97)	3,09 (1,19)	2,21 (1,06)	2,50 (1,03)	3,64 (1,23)	2,09 (1,02)
Post-hoc (Bonferroni)							
Faktor	(1) vs. (3)**	(1) vs. (3)**	(1) vs. (3)**	(1) vs. (2)**	(1) vs. (2)*	(1) vs. (3)**	(1) vs. (2)**
Statusgruppe	(2) vs. (3)**	(2) vs. (3)**		(2) vs. (3)**	(2) vs. (3)**	(2) vs. (3)**	(2) vs. (3)**

Bem.: M: Mittelwert; SD: Standardabweichung; *: $p \leq 0,05$; **: $p \leq 0,01$.

Über die insgesamt sieben Erklärungen liegen 13 signifikante Unterschiede zwischen einzelnen Statusgruppen vor. Mit Ausnahme von Video TA_a unterscheiden sich alle Bewertungen zwischen Studierenden und Lehrenden signifikant. Die dabei nach der Konvention von Cohen (1988, Kap. 2) mittleren bis großen vorliegenden Effekte sind Tabelle 5.19 zu entnehmen.

Tabelle 5.19 Effektstärken zu den Mittelwertsunterschieden zwischen **Studierenden und Lehrenden**

Video	Erklärthema	Erklärmodus	t	df	p	d
	Notenwerte	optisch	-3,33	108	< 0,01	-0,61
	Notenwerte	optisch & akustisch	-4,57	108	< 0,01	-0,82
	Taktarten	akustisch	-1,64	116	n. s.	-0,30
	Taktarten	optisch & akustisch	-3,55	117	< 0,01	-0,65
	Tonleiter	optisch	-5,20	95	< 0,01	-0,95
	Tonleiter	akustisch	-3,99	118	< 0,01	-0,73
	Tonleiter	optisch & akustisch	-3,87	116	< 0,01	-0,71

Bem.: t: T-Wert; df: Freiheitsgrade; p: Überschreitungswahrscheinlichkeit; d: Effektstärke nach Cohen.

Diese klaren Ergebnisse zeigen in sechs der sieben Videos signifikante Unterschiede in den absoluten Bewertungen zwischen den beiden Gruppen. Auch sehen die Lehrenden mit Blick auf den Bereich der Bewertungsskala noch mehr Potenzial zur Verbesserung, um von einer *sehr guten* Erklärung zu sprechen. Dennoch müssen die Ergebnisse anhand der grafischen Analyse, die annähernd parallele Bewertungstendenzen deutlich werden lässt, eingeschränkt werden (siehe Abb. 5.8). Demnach kann *Hypothese 3.2*, dass Studierende und Lehrende die Qualität von Erklärungen unterschiedlich bewerten, aus statistischer Perspektive aufrechterhalten werden. Unter Berücksichtigung der ähnlichen Bewertungstendenzen wird deskriptiv betrachtet jedoch nicht von einer grundsätzlich divergierenden Vorstellung beider Gruppen zu Erklärqualität ausgegangen.

Hinsichtlich potenzieller Unterschiede zwischen den beiden Perspektiven auf unterrichtliches Erklären durch Schüler*innen auf der einen und Studierende sowie Lehrende auf der anderen Seite zeigen die Ergebnisse in Tabelle 5.18 signifikante Unterschiede in allen sieben Erklärungen: Vier signifikante Unterschiede liegen zwischen Schüler*innen und Lehrenden bei den Erklärungen NW_o, NW_oa, TA_a und TL_a vor, die entweder kleine (NW_o; TA_a) oder große Effekte (NW_oa; TL_a) zeigen (siehe Tabelle 5.20).

Tabelle 5.20 Effektstärken zu den Mittelwertsunterschieden zwischen **Schüler*innen und Lehrenden**

Video					
Erklärthema	Erklärmodus	t	df	p	d
Notenwerte	optisch	-2,88	203	< 0,01	-0,44
Notenwerte	optisch & akustisch	-5,56	205	< 0,01	-0,84
Taktarten	akustisch	-3,13	202	< 0,01	-0,48
Taktarten	optisch & akustisch	0,74	202	n. s.	0,11
Tonleiter	optisch	-1,98	202	n. s.	-0,31
Tonleiter	akustisch	-5,51	203	< 0,01	-0,85
Tonleiter	optisch & akustisch	0,20	202	n. s.	-0,48

Bem.: t: T-Wert; df: Freiheitsgrade; p: Überschreitungswahrscheinlichkeit; d: Effektstärke nach Cohen.

Die *Post-hoc*-Tests ergeben zudem drei signifikante Unterschiede zwischen den Gruppen der Schüler*innen und den Studierenden, denen kleine und mittlere Effekte zugrunde liegen (siehe Tab. 5.21). Damit kann auch *Hypothese 3.1*, dass sich die Bewertungen der Schüler*innen von denen der Studierenden und Lehrenden signifikant unterscheiden, unterstützt werden. Mit Blick auf Abbildung 5.8 zeigt sich dabei auch, dass keine ähnlichen Bewertungstendenzen zwischen Schüler*innen sowie einer der beiden anderen Gruppen vorliegen. Die Annahme einer perspektivenspezifischen Validität von Erklärqualität – im wörtlichen Sinn: der zwei Perspektiven auf Erklären – kann damit aufrechterhalten werden.

Tabelle 5.21 Effektstärken zu den Mittelwertsunterschieden zwischen **Schüler*innen und Studierenden**

Video					
Erklärthema	Erklärmodus	T	df	p	d
Notenwerte	optisch	0,56	203	n. s.	0,09
Notenwerte	optisch & akustisch	-0,45	204	n. s.	-0,07
Taktarten	akustisch	-1,23	202	n. s.	-0,19
Taktarten	optisch & akustisch	5,16	159	< 0,01	0,73
Tonleiter	optisch	3,50	194	< 0,01	0,47
Tonleiter	akustisch	-1,32	203	n. s.	-0,21
Tonleiter	optisch & akustisch	5,18	164	< 0,01	0,73

Bem.: t: T-Wert; df: Freiheitsgrade; p: Überschreitungswahrscheinlichkeit; d: Effektstärke nach Cohen.

Zusammenfassung

Zusammenfassend können auf die dritte Forschungsfrage, wie sich die Globalurteile zur Erklärqualität zwischen Schüler*innen, Studierenden und Lehrenden unterscheiden, folgende Antworten gegeben werden:

Es liegen signifikante Unterschiede in den durchschnittlichen Bewertungen der drei Statusgruppen vor. Die Gruppe der Schüler*innen zeigt dabei ein anderes Bewertungsverhalten als Studierende und Lehrkräfte. Zum einen vergeben sie ihre Globalurteile auf einem kleineren Bereich der sechsstufigen Notenskala, zum anderen bewerten sie insbesondere die beiden Erklärungen zu Notenwerten besser. Da die Selbsteinschätzungen der Schüler*innen hinsichtlich ihres eigenen Vorwissens zum Thema Notenwerte lediglich auf Itemebene erhoben wurden, können sie inferenzstatistisch nicht ausgewertet werden. Allerdings fallen die Selbsteinschätzungen deskriptiv betrachtet niedriger aus. Das bedeutet, dass Schüler*innen glauben, mehr zum Thema Notenwerte zu wissen als zu den anderen beiden Themen. Diese Einschätzungen bestätigen Ergebnisse des kurzen Wissenstests (siehe Kap. 5.5), die ein höheres Vorwissen der Schüler*innen zum Thema Notenwerte feststellen.

Die Selbsteinschätzung der Schüler*innen zur Frage, wie sie das Erklärte verstanden haben, korrelieren hoch mit den Globalurteilen. Es kann daher angenommen werden, dass ein höheres Vorwissen sowie die Vermutung eines damit einhergehenden Gefühls, die Erklärungen auch besser verstanden zu haben, die Bewertung der Erklärqualität durch die Schüler*innen beeinflusst. Auch bei Studierenden und Lehrenden korrelieren die eigenen Globalbewertungen zur Erklärqualität hoch mit der Einschätzung zur Frage, inwieweit Schüler*innen die Inhalte der Erklärungen verstehen. Allerdings unterscheiden sich die Urteile dieser beiden Gruppen von denen der Schüler*innen insbesondere darin, dass Studierende und Lehrende die höchste Qualität den Erklärungen TA_oa und TL_oa zuschreiben und auch glauben, dass Schüler*innen einer sechsten Jahrgangsstufe diese zwei Erklärungen am besten verstehen.

Wie bereits dargelegt limitieren die Interaktionseffekte zwischen Statusgruppe und Erklärmodus bei den Themen Taktarten und Tonleitern zudem die Aussagekraft des Faktors Statusgruppe innerhalb der jeweiligen Varianzanalyse, da dieser durch die Ausprägungen des anderen Faktors Erklärmodus bedingt ist. Dennoch zeigt sich, dass die Unterscheidung der Statusgruppen hilfreiche Erkenntnisse liefert, anhand derer *Hypothese 3.1* und *Hypothese 3.2* unterstützt werden können. Eine detaillierte weiterführende Analyse des Faktors Erklärmodus erfolgt im nächsten Kapitel.

5.3.3 Faktor Erklärmodus

F4) Wie unterscheiden sich die Bewertungen von Erklärqualität in Abhängigkeit des Erklärmodus?

In der eingangs durchgeführten Varianzanalyse mit gemischtem Design (siehe Kap. 5.3.1) zeigte sich unter Berücksichtigung von Interaktionseffekten der beiden Faktoren im Modell, dass weiterführende Analysen – zumindest für die Erklärungen zu Taktarten und Tonleiter – hinsichtlich des Erklärmodus ohne Einschränkungen durchführbar und sinnvoll sind.

Die Darstellung der Ergebnisse erfolgt im Top-Down-Prinzip: Zuerst werden die drei Erklärmodi auf Basis der durchschnittlichen Bewertungen aller Studienteilnehmer*innen miteinander verglichen. In einem zweiten Schritt werden beide Repräsentationsformen fokussiert betrachtet. Abschließend werden die Begründungen zur Bewertung der Erklärqualität von Studierenden und Lehrenden in Form offener Antworten einbezogen. Da diese im ersten Durchgang direkt im Anschluss an die jeweiligen Erklärungen gegeben wurden, waren die Studienteilnehmer*innen nicht durch konkrete Items zum Einsatz musikbezogener Repräsentationen beeinflusst.¹²⁶ Relevant für die Analyse sind insbesondere solche Textstellen, die generelle Überzeugungen der Erklärenden hinsichtlich des Einsatzes von Repräsentationen beim Erklären zum Vorschein bringen (siehe *Kategoriensystem 2b, Anhang F4*).

Die Erklärmodi im Vergleich

Vor dem Hintergrund der Frage nach Unterschieden in der durchschnittlichen Globalbewertung der Erklärqualität, die sich auf die verschiedenen Erklärmodi zurückführen lassen, wurde zunächst folgende Erwartung formuliert: Erklärungen, die beide Repräsentationsformen miteinander kombinieren, werden besser bewertet als solche, in denen nur eine Form der Repräsentation zum Einsatz kommt.

Dazu ist in Tabelle 5.22 zunächst deskriptiv festzustellen, dass unabhängig vom Thema jeweils die Erklärungen von allen Studienteilnehmer*innen am besten bewertet werden, die sowohl eine optische als auch eine akustische Repräsentation zum Einsatz bringen (NW_oa: 1,83; TA_oa: 2,12; TL_oa: 1,95). Mit Ausnahme der nahezu gleich ausfallenden Globalbewertung der Lehrenden für beide Erklärungen zum Thema Notenwerte zeigt sich dieses Bild auch unabhängig vom Thema mit Blick auf alle Statusgruppen: Zwar fallen die jeweiligen Abstände zu den übrigen Videos unterschiedlich aus, jedoch werden pro Thema nahezu immer die Erklärungen mit kombiniertem Erklärmodus (optische und akustische Repräsentationsform) am besten bewertet. Die beiden Erklärungen, in denen eine optische Repräsentation fehlt (TA_a: 2,68; TL_a: 2,91), erhalten nicht

¹²⁶ Die Auswertungen in Abschnitt 5.3.3 beziehen sich hinsichtlich des Einsatzes optischer und akustischer Repräsentationsformen ausschließlich auf die systematische Variation auf Videoebene und nicht auf die Skala *Einsatz musikbezogener Repräsentationen*.

nur insgesamt, sondern auch von allen drei Statusgruppen jeweils die schlechtesten Bewertungen (siehe Tab. 5.22). Dabei sprechen alle Gruppen der entsprechenden Erklärung zu Tonleitern (TL_a) im Vergleich die geringste Qualität zu. Eine Art Mittelposition in der Gesamtschau nehmen die Erklärungen mit ausschließlich optischer Repräsentation ein (NW_o; TL_o).

Tabelle 5.22 Mittelwerte (Standardabweichungen) der Globalurteile pro Video und Statusgruppe

Erklärthema Erklärmodus Videokürzel	Notenwerte		Taktarten		Tonleiter		
	opt.	opt. & aku.	aku.	opt. & aku.	opt.	aku.	opt. & aku.
	NW_o	NW_oa	TA_a	TA_oa	TL_o	TL_a	TL_oa
Statusgruppe	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Schüler*innen ($N_{max} = 148$)	1,90 (1,01)	1,65 (0,81)	2,51 (1,10)	2,32 (1,21)	2,15 (1,22)	2,61 (1,27)	2,10 (1,07)
Studierende ($N_{max} = 61$)	1,82 (0,72)	1,70 (0,63)	2,71 (1,05)	1,56 (0,83)	1,70 (0,63)	2,86 (1,02)	1,45 (0,69)
Lehrende ($N_{max} = 62$)	2,35 (0,97)	2,36 (0,97)	3,09 (1,19)	2,21 (1,06)	2,50 (1,03)	3,64 (1,23)	2,09 (1,02)
Globalurteile							
($N_{max} = 271$)	1,99 (0,96)	1,83 (0,88)	2,68 (1,13)	2,12 (1,14)	2,13 (1,10)	2,91 (1,27)	1,95 (1,02)

Bem.: *M*: Mittelwert; *SD*: Standardabweichung.

Analysiert man die Unterschiede aller sieben Globalurteile hinsichtlich der drei systematisch variierten Erklärmodi inferenzstatistisch – in Kapitel 5.3.1 wurde der Faktor Erklärmodus jeweils pro Erklärthema und nicht über alle Erklärungen hinweg untersucht –, zeigen sich signifikante Unterschiede ($F(2, 1831) = 71.01; p < 0,01; \eta^2 = 0,07$) mit einem mittleren Effekt.

Bonferroni-korrigierte *Post-hoc*-Tests präzisieren die Beobachtung: Signifikante Unterschiede liegen nur zwischen dem Erklärmodus *akustisch* und *optisch* ($p < 0,01$) beziehungsweise *akustisch* und *optisch und akustisch* ($p < 0,01$), nicht jedoch zwischen *optisch* und *optisch und akustisch* ($p = 0,41$) vor. Das bedeutet, dass das Fehlen optischer Repräsentation mit deutlich schlechteren Bewertungen der Erklärqualität einhergeht als der Verzicht auf akustische Repräsentation – zumindest in der Form von Hör- und Klangbeispielen, die bei Erklärungen der vorliegenden Studie verwendet wurden.

Um im Folgenden die Unterschiede zwischen den einzelnen Erklärmodi detailliert beziehungsweise auch unter Berücksichtigung der Statusgruppen zu untersuchen, werden *Post-hoc*-Analysen durchgeführt. Da der Faktor Erklärmodus beim Thema Tonleiter im gemischten Modell dreistufig berücksichtigt wurde – bei Notenwerten und Taktarten liegt der Faktor zweistufig vor –, erfolgt vorbereitend eine Analyse dieses Faktors. Tabelle 5.23 zeigt im Folgenden, dass sich die Globalurteile von TL_o (optische Repräsentation) und TL_oa (optische und akustische Repräsentation) nicht signifikant unterscheiden ($t(261) = 2,60; n. s.; d = 0,16$). Im Gegensatz dazu liegen signifikante Unterschiede zwischen TL_a und TL_o ($t(261) = -9,33; p < 0,01; d = -0,58$) sowie TL_oa ($t(261) = 11,17; p < 0,01; d = 0,69$) vor, die jeweils mittlere Effektstärken aufweisen und für weitere Analysen berücksichtigt werden können.

Tabelle 5.23 Einfaktorielle ANOVA der drei Erklärungen zum Thema Tonleiter mit *Bonferroni*-Korrektur

Erklärthema Erklärmodus Videokürzel	Tonleiter		
	(1) opt.	(2) aku.	(3) opt. & aku.
Globalurteile	TL_o	TL_a	TL_oa
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
(<i>N</i> _{max} = 271)	2,13 (1,10)	2,91 (1,27)	1,95 (1,02)
Post-hoc (<i>Bonferroni</i>)			
Faktor	(1) vs. (2)**		
Erklärmodus	(2) vs. (3)**		

Bem.: *M*: Mittelwert; *SD*: Standardabweichung; *: $p \leq 0,05$; **: $p \leq 0,01$.

Akustische Repräsentation

Nun richtet sich der Blick auf die Unterschiede zwischen Erklärungen mit optischen und solchen mit optischen und akustischen Repräsentationen unter Berücksichtigung der einzelnen Statusgruppen. Anhand dessen wird die Frage beantwortet, welche Relevanz akustischen Repräsentationen zugeschrieben wird. Da im gemischten Modell (siehe Kap. 5.3.1) auch der Faktor Statusgruppe einbezogen wurde, soll im Folgenden analysiert werden, ob Unterschiede hinsichtlich des Erklärmodus auch zwischen den Statusgruppen zu ermitteln sind. Um die jeweiligen Effektstärken mit den später folgenden Unterschieden der Erklärungen mit akustischen Repräsentationen deutlich zu machen, werden *t*-Tests durchgeführt, obwohl weder im Modell zum Thema Notenwerte (siehe Kap. 5.3.1) noch innerhalb der mittels *Bonferroni*-Korrektur (siehe Tab. 5.24) durchgeführten *Post-hoc*-Tests signifikante Unterschiede vorliegen.

Tabelle 5.24 Unterschiede zwischen Erklärungen mit optischer Repräsentation und der Kombination aus optischer und akustischer Repräsentation unter Berücksichtigung der Statusgruppen

Video		<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Erklärthema	Statusgruppe				
Notenwerte	Schüler*innen	2,73	144	< 0,01	0,23
NW_o	Studierende	1,07	59	n. S.	0,14
NW_oa	Lehrende	-0,09	59	n. S.	-0,01
Tonleiter	Schüler*innen	0,51	144	n. S.	0,04
TL_o	Studierende	2,38	58	0,02	0,31
TL_oa	Lehrende	3,34	57	< 0,01	0,44

Bem.: *t*: *T*-Wert; *df*: Freiheitsgrade; *p*: Überschreitungswahrscheinlichkeit, *d*: Effektstärke nach Cohen.

Aus Tabelle 5.24 geht hervor, dass die Effekte für das Fehlen einer akustischen Repräsentation entweder nicht signifikant sind oder nur klein ausfallen. Es lässt sich dabei unabhängig von Erklärthema und Zugehörigkeit zur Statusgruppe keine Systematik erkennen. Das bedeutet, dass der Einsatz akustischer Repräsentationen, so wie sie in den

Erklärungen der vorliegenden Studie umgesetzt wurden, keinen bedeutenden Effekt auf die Bewertung der Erklärqualität hat. Das ist zumindest dahingehend überraschend, als die klangliche Darstellung musikalischer Phänomene ein fachspezifisch entscheidendes Prinzip guten Musikunterrichts zu sein scheint.

Optische Repräsentation

Im Gegensatz dazu fallen die Unterschiede zwischen Erklärungen, die nur akustische oder optische und akustische Repräsentation einsetzen, deutlicher aus. Dadurch sind Aussagen über die Bedeutung optischer Repräsentationen möglich. Sowohl bei den Unterschieden zwischen den durchschnittlichen Globalurteilen beider Taktarten-Erklärungen (TA_a und TA_oa; siehe Kap. 5.3.1) als auch der relevanten Tonleiter-Erklärungen (TL_a und TL_oa; siehe Tabelle 5.23) können Unterschiede auf den Faktor Erklärmodus zurückgeführt werden.

Tabelle 5.25 *Post-hoc*-Tests mit *Bonferroni*-Korrektur zu den Erklärungen mit akustischer Repräsentation und der Kombination aus optischer und akustischer Repräsentation

Erklärthema Erklärmodus Videokürzel	Taktarten				Tonleiter			
	akustisch		opt. & aku.		akustisch		opt. & aku.	
Statusgruppe	TA_a	TA_oa	TA_a	TA_oa	TL_a	TL_oa	TL_a	TL_oa
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Schüler*innen (<i>N_{max}</i> = 148)	2,51 (1,10)	2,32 (1,21)	2,61 (1,27)	2,10 (1,07)				
Studierende (<i>N_{max}</i> = 61)	2,71 (1,05)	1,56 (0,83)	2,86 (1,02)	1,45 (0,69)				
Lehrende (<i>N_{max}</i> = 62)	3,09 (1,19)	2,21 (1,06)	3,64 (1,23)	2,09 (1,02)				
Globalurteile (<i>N_{max}</i> = 271)	2,68 (1,13)	2,12 (1,14)	2,91 (1,27)	1,95 (1,02)				
Post-hoc (<i>Bonferroni</i>)	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Faktor Erklärmodus	7,04	261	< 0,01	0,43	11,17	261	< 0,01	0,69

Bem.: *M*: Mittelwert; *SD*: Standardabweichung; *t*: T-Wert; *df*: Freiheitsgrade; *p*: Überschreitungswahrscheinlichkeit; *d*: Effektstärke nach Cohen.

Aus Tabelle 5.25 gehen die Ergebnisse von *Post-hoc*-Tests über die Gesamtstichprobe hervor. Unabhängig von den Themen Taktarten und Tonleiter fallen die Globalurteile der Erklärungen, in denen eine optische Repräsentation fehlt, deutlich schlechter aus. Dies zeigt sich tendenziell auch, wenn die drei Statusgruppen einzeln betrachtet werden (siehe Tab. 5.26).

Tabelle 5.26 Unterschiede zwischen Erklärungen mit akustischer Repräsentation und der Kombination aus optischer und akustischer Repräsentation unter Berücksichtigung der Statusgruppen

Video		<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Erklärthema	Statusgruppe				
Taktarten	Schüler*innen	1,73	144	n. s.	0,14
TA_a	Studierende	8,74	58	< 0,01	1,14
TA_oa	Lehrende	5,58	57	< 0,01	0,75
Tonleiter	Schüler*innen	4,69	144	< 0,01	0,39
TL_a	Studierende	9,78	57	< 0,01	1,27
TL_oa	Lehrende	8,60	57	< 0,01	1,13

Bem.: *t*: *T*-Wert; *df*: Freiheitsgrade; *p*: Überschreitungswahrscheinlichkeit, *d*: Effektstärke nach Cohen.

Aus Tabelle 5.26 geht hervor, dass die Unterschiede bei der Gruppe der Schüler*innen entweder nicht signifikant (TA_a, TA_oa) oder signifikant, aber nur mit kleinen Effekten (TL_a, TL_oa), ausfallen; im Gegensatz dazu zeigen sich bei Studierenden (Thema Taktarten: $d = 1,14$; Thema Tonleiter: $d = 1,27$) durchweg sehr große und bei der Gruppe der Lehrenden mittlere (Thema Taktarten: $d = 0,75$) und sehr große Effekte (Thema Tonleiter: $d = 1,13$). Diese sind im Vergleich zu den Effektstärken, die in Tabelle 5.24 hinsichtlich des Einsatzes akustischer Repräsentationen präsentiert wurden, deutlich stärker.

Eine weitere Beobachtung stützt diese Ergebnisse. Bei allen Erklärungen, in denen auf eine Repräsentationsform verzichtet wurde (NW_o, TA_a, TL_o, TL_a), sind folgende Items gestellt worden: Einerseits wurden die Schüler*innen gefragt, ob sie sich das Vorkommen einer Präsentation an der Tafel (optische Repräsentation) oder eines Hörbeispiels (akustische Repräsentation) gewünscht hätten; andererseits wurde an Studierende und Lehrende die Frage gerichtet, ob sich Schüler*innen einer sechsten Jahrgangsstufe das eben Genannte wünschen würden. Tabelle 5.27 gibt eine Übersicht der sechsstufigen Skalen, bei denen der Wert 1 die größtmögliche Zustimmung und der Wert 6 die größtmögliche Ablehnung darstellen.

Tabelle 5.27 Deskriptive Ergebnisse pro Statusgruppe zum Wunsch nach akustischer beziehungsweise optischer Repräsentation

Erklärthema	Wunsch akustisch		Wunsch optisch	
	optisch	optisch	akustisch	akustisch
Erklärmodus	NW_o	TA_o	TA_a	TL_a
Videokürzel				
Statusgruppe	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Schüler*innen	2,12	2,19	1,99	2,10
($N_{max} = 148$)	(1,44)	(1,41)	(1,49)	(1,51)
Studierende	2,21	1,91	1,74	1,38
($N_{max} = 61$)	(1,10)	(0,96)	(1,06)	(0,87)
Lehrende	1,76	1,59	1,91	1,40
($N_{max} = 62$)	(1,10)	(0,89)	(1,06)	(0,77)
Alle Gruppen				
($N_{max} = 271$)	2,07	2,00	1,91	1,80
	(1,31)	(1,24)	(1,33)	(1,31)

Bem.: *M*: Mittelwert; *SD*: Standardabweichung.

Die Auswertung erfolgt rein deskriptiv, da diese Angaben lediglich als Kontrollvariable auf Ebene von Einzelitems erhoben wurden und somit eine geringe Zuverlässigkeit der Aussagen erwartet wird. Zunächst ist festzuhalten, dass der Wunsch sowohl nach optischen als auch nach akustischen Repräsentationen insgesamt sowie auch in den einzelnen Gruppen hoch ist. Dabei zeigt sich jedoch mit Ausnahme der Einschätzung der Lehrenden bei der Erklärung TA_a eine generell höhere Zustimmung bei dem Wunsch nach optischer Repräsentation, wenn diese in einer Erklärung nicht vorgekommen ist. Das deckt sich mit den deutlich stärkeren Effekten hinsichtlich der Unterschiede in den durchschnittlichen Globalurteilen zwischen den Erklärungen mit akustischer Repräsentation und solchen, die eine Kombination aus optischer und akustischer Repräsentation enthalten (siehe Tab. 5.26). Damit fällt das Fehlen einer optischen Repräsentation negativer ins Gewicht.

Bei der Gruppe der Schüler*innen sticht zweierlei hervor: Zum einen fallen die Einschätzungen uneinheitlich aus, was aus den hohen Standardabweichungen hervorgeht (Min: 1,41; Max: 1,51). Es wird vermutet, dass hier Unterschiede im Vorwissen der Schüler*innen zu diesen hohen Werten beitragen können – in Kapitel 5.5 wird diese Annahme überprüft. Zum anderen fallen die Unterschiede hinsichtlich der Wünsche nach beiden Repräsentationsformen im Gegensatz zu den anderen beiden Gruppen ähnlich marginal aus wie die Unterschiede in den Globalbewertungen der Erklärqualität.

Deutlicher unterscheiden die Studierenden in ihren Einschätzungen bezüglich des Wunsches von Schüler*innen nach optischer und akustischer Repräsentation. Anscheinend gehen Studierende davon aus, dass Schüler*innen optische Repräsentationen beim Erklären für notwendiger befinden als akustische Repräsentationen. Bei der Gruppe der Lehrenden sind ebenfalls zwei Aspekte auffallend: Zum einen fällt ihre Zustimmung auf die Frage, ob sich Schüler*innen akustische Repräsentationen wünschen, höher aus als in den anderen beiden Gruppen. Zum anderen halten Lehrende – ebenso wie Studierende – insbesondere bei der Erklärung TL_a, der die schlechteste Qualität attestiert wurde, eine optische Repräsentation für zwingend erforderlich.

Zwischenfazit I

Bezüglich des Einsatzes von Repräsentationen ist folgendes erstes Fazit festzuhalten: Mit Blick auf die Gesamtstichprobe werden Erklärungen, die sowohl optische wie akustische Repräsentationen zum Einsatz bringen, am besten bewertet. Damit kann *Hypothese 4.1* aufrechterhalten werden, die für eine Kombination aus optischen und akustischen Repräsentationen die besten Bewertungen der Erklärqualität erwartet. Weiterhin werden auch Erklärungen, bei denen eine akustische Repräsentation fehlt, nicht signifikant schlechter bewertet. Deutlich negativer und signifikant schlechter fallen die Bewertungen aus, wenn eine optische Repräsentation ausbleibt.

Diese Reihenfolge zeigt sich auch beim gesonderten Blick auf die einzelnen drei Statusgruppen: Festzustellen sind ähnliche Bewertungstendenzen hinsichtlich der zum Einsatz kommenden Erklärmodi, auch wenn teilweise unterschiedliche Bereiche der Notenskala verwendet werden oder Unterschiede zwischen verschiedenen Erklärungen zum

selben Thema unterschiedlich stark ausfallen. Forschungsfrage F4b ist damit zugunsten des Einsatzes optischer Repräsentationen zu beantworten, da diese Erklärungen besser bewertet werden. Im Folgenden werden die Perspektiven der Studierenden und Lehrenden weiter ausdifferenziert, indem abschließend deren Begründungen zu ihrer Bewertung der Erklärqualität in Form offener Antworten analysiert werden.

Überzeugungen zum Einsatz akustischer Repräsentationen

Äußerungen zum Einsatz akustischer Repräsentationen treten in den offenen Antworten vorwiegend bei den Erklärungen auf, in denen solche fehlen. So finden sich etwa die Hälfte der codierten Stellen in den beiden Videos ohne akustische Repräsentation (NW_o, TL_o). In den Begründungen zur Bewertung der Erklärqualität finden sich beispielsweise Aussagen wie „die akustische Ebene wird komplett ausgeblendet“ (*Prof.1_TL.o_Bg*), oder „Klangeindruck fehlt, obwohl dieser sehr wichtig ist!“ (*Stud.58_NW.o_Bg*). An weiteren Stellen wird in knapper Form auf das Fehlen einer akustischen Repräsentation hingewiesen (z. B. *Slk.15_NW.o_Bg*, *Slk.7_TL.o_Bg*, *Stud.9_TL.o_Bg*). Die andere Hälfte der Codierungen erfolgt bei den Erklärungen, die eine klangliche Form der Repräsentation aufweisen (NW_oa, TA_a, TA_oa, TL_a, TL_oa). Der Einsatz von akustischen Repräsentationen wird bei den Erklärungen, die den kombinierten Erklärmodus verwenden (NW_oa, TA_oa, TL_oa), als Form der Unterstützung hervorgehoben (z. B. *Stud.60_NW.oa_Bg*; *Slk.15_NW.oa_Bg*; *Hsl.8_TA.oa_Bg*; *Lk.19_TL.oa_Bg*).

Nicht selten werden dabei auch Erklärungen zu demselben Thema miteinander verglichen. Dabei bewertet eine Seminarlehrkraft die Qualität der Erklärung in Video TL_a „wesentlich besser, da anhand des Höreindrucks erklärt wird“, wobei sie zugleich einschränkt, dass die visuelle Unterstützung fehle (*Slk.6_TL.a_Bg*).¹²⁷ Daher kann angenommen werden, dass beide Repräsentationsformen wichtig sind und eine Kombination von der Seminarlehrkraft favorisiert wird. Hinweise auf eine Gewichtung beider Repräsentationsformen finden sich zum Beispiel dahingehend, dass eine klangliche Darstellung ohne Visualisierung nicht sinnvoll sei (*Hsl.9_TL.a_Bg*). Das stützt die bisherigen Beobachtungen zu besseren Bewertungen von Erklärungen, in denen eine optische Repräsentation zum Einsatz kommt.

Weitere Aussagen wurden auf die Bitte nach Vorschlägen zur Verbesserung der Erklärung von Studierenden wie Lehrenden getätigt. Textstellen, die den Einsatz akustischer Repräsentationen thematisieren, sind nahezu ausschließlich bei den Erklärungen zu finden, die auf eine akustische Repräsentation verzichten. Empfohlen wird beispielsweise bei der entsprechenden Erklärung zur Tonleiter (TL_o), diese vorzuspielen oder vorzusingen (z. B. *Hsl.8_TL.o_Vb*; *Slk.5_TL.o_Vb*; *Stud.31_TL.o_Vb*). Auch für die Erklärung zu Notenwerten, in der keine akustische Repräsentation gegeben wird (NW_o), wird die Verwendung eines Hörbeispiels empfohlen (z. B. *Hsl.9_NW.o_Vb*; *Stud.24_NW.o_Vb*).

¹²⁷ Der Komparativ resultiert vermutlich daher, weil zuvor ein anderes Video (TL_o) gesehen wurde, welches keine akustische Repräsentation verwendet. Dort bemängelt die Seminarlehrkraft allerdings, dass eine akustische Darstellung fehlt.

Insgesamt geht aus den offenen Antworten eine hohe Bedeutung des Einsatzes akustischer Repräsentationen hervor. Allerdings werden auch Einschränkungen dahingehend vorgenommen, dass eine akustische Repräsentation bei gleichzeitigem Fehlen einer optischen nicht hinreichend sein kann. Auch wenn kein eindeutiges Bild hinsichtlich des Einsatzes akustischer Repräsentationen erkennbar ist, überwiegen hinsichtlich der Qualität von Erklärungen die positiv konnotierten Aussagen.

Überzeugungen zum Einsatz optischer Repräsentationen

Im Vergleich zu akustischen Repräsentationen lassen sich ähnlich viele Äußerungen hinsichtlich des Fehlens beziehungsweise Vorkommens optischer Repräsentationen finden. Sie werden nahezu ausschließlich bei den Erklärungen TA_a und TL_a abgegeben – also auch hier bei den Erklärungen, in denen eine optische Repräsentation fehlt. Die Aussagen variieren zwischen „fehlende Visualisierung“ (*Lk.6_TL.a_Bg*), „keinerlei optische Veranschaulichung“ (*Slk.5_TL.a_Bg*) oder „visuelle Erklärung fehlt komplett“ (*Stud.7_TL.a_Bg*). An manchen Stellen werden auch Vergleiche mit anderen Erklärungen zu demselben Thema gezogen, die eine optische Repräsentation beinhalten. So wurde das Thema Tonleiter „mit Visualisierung besser erklärt“ (*Hsl.12_TL.a_Bg*) oder im Vergleich zu den beiden anderen Erklärungen (TL_o, TL_oa) das fehlende Anschauungsmaterial kritisiert (z. B. *Stud.17_TL.a_Bg*), was als Grund für ein schlechteres Urteil genannt wurde.

Während an vielen Stellen nicht näher spezifiziert wird, was mit Visualisierung oder Veranschaulichung explizit gemeint ist, wird in manchen Äußerungen konkret bemängelt, dass die „Visualisierung der Töne“ (*Hsl.3_TL.a_Bg*), der „Notentext“ (*Stud.30_TA.a_Bg*), die „grafische Darstellung der Notenwerte“ (*Stud.14_TA.a_Bg*), die „Notation“ (*Stud.30_TA.a_Bg*) oder auch das „Notationsbild“ (*Stud.38_TA.a_Bg*) fehlen. Fernab des Notenbildes werden selten weitere konkrete Beispiele für Visualisierungen genannt. Vereinzelt wird beispielsweise die grafische Darstellung der Klaviatur gefordert, um Ganz- und Halbtonschritte der Tonleiter zu veranschaulichen (z. B. *Lk.21_TA.a_Bg*; *Lk.13_TL.a_Bg*; *Hsl.9_TL.a_Bg*).

Der Wunsch nach dem Einsatz optischer Repräsentationen wird insbesondere bei den Erklärungen geäußert, in denen eine solche fehlt (TA_a, TL_a). Dabei zeigt sich dasselbe Muster wie beim Wunsch nach akustischen Repräsentationen, die vor allem im Kontext der Erklärungen NW_o und TL_o gefordert wurden: Konkrete Vorschläge beziehen sich auf den Einsatz von Dokumentenkameras (z. B. *Prof.6_TL.a_Vb*; *Slk.5_TL.a_Vb*) oder die Tafel (z. B. *Stud.5_TL.a_Vb*; *Hsl.4_TA.a_Vb*; *Slk.4_TA.a_Vb*). An manchen Textstellen wird der Einsatz optischer Repräsentationen explizit mit der musikalischen Vorerfahrung der Schüler*innen verknüpft: So seien Erklärungen, in denen eine optische Repräsentation fehlt, insbesondere von Schüler*innen mit fehlender musikalischer Vorerfahrung schwer zu verstehen (z. B. *Stud.26_TL.a_Bg*; *Stud.35_TL.a_Bg*; *Lk.23_TL.a_Bg*).

Darüber hinaus finden sich Textstellen zum Einsatz optischer Repräsentationen, die absolute Aussagen wie „die beste Erklärung hilft in diesem Fall ohne den visuellen Zugang nichts“ (*Lk.23_TA.a_Rf*), „Visualisierungen sind unabdingbar“ (*Slk.5_TL.a_Vb*)

oder „für Sechstklässler [*sind*, vom Autor hinzugefügt] Visualisierungen dringend notwendig“ (*Lk.23_TL.a_Bg*) beinhalten. Dabei fällt auf, dass solche absoluten Aussagen nur im Kontext des Einsatzes optischer und nicht im Kontext akustischer Repräsentationen zu finden sind. Dies kann als weiteres Indiz dafür gewertet werden, dass dem Einsatz optischer Repräsentationen eine bedeutendere Rolle hinsichtlich der Qualität unterrichtlichen Erklärens im Fach Musik zuteilwird als akustischen Repräsentationen – zumindest, wenn sie von den Schüler*innen, wie in den Erklärsequenzen rezipiert, und beispielsweise nicht durch eigenes Singen oder Musizieren umgesetzt werden.

Zwischenfazit II

Vergleicht man die bisherigen Aussagen zu optischen und akustischen Repräsentationen, können sowohl Ähnlichkeiten als auch Unterschiede festgestellt werden. Gemeinsam ist beiden Repräsentationsformen, dass sich Textstellen insbesondere bei den Erklärungen finden lassen, die auf eine der beiden Repräsentationsformen verzichten: In den Begründungen zur Bewertung der Erklärqualität erwähnen die Studienteilnehmer*innen bei den Erklärungen NW_o und TL_o häufig das Fehlen akustischer, in den Erklärungen TA_a und TL_a vor allem das Ausbleiben optischer Repräsentationen.

Unterschiede lassen sich hingegen tendenziell in der Häufigkeit der Nennung als auch in der Intensität der Begründung erkennen. Diese decken sich mit den Bewertungen zur Erklärqualität, die zu jedem Video als Globalurteile abgegeben wurden (siehe Tab. 5.14): Der Einsatz optischer Repräsentationen wird nicht nur häufiger thematisiert, auch scheint er aus Sicht von Studierenden und Lehrenden dringend notwendig oder gar unabdingbar (z. B. *Lk.23_TL.a_Bg*; *Slk.5_TL.a_Vb*), damit Schüler*innen einer sechsten Jahrgangsstufe das Erklärte verstehen können. Diese Beobachtung deckt sich mit den signifikant schlechter ausfallenden Globalurteilen der Erklärungen TA_a und TL_a.

Damit stützen sich die empirischen Befunde der quantitativen wie qualitativen Analyse gegenseitig, sodass weiterhin angenommen werden kann, dass der Einsatz optischer Repräsentationen beim Erklären musiktheoretischer Inhalte zu einer höheren Erklärqualität führe als der Einsatz akustischer Repräsentationen (siehe Forschungsfrage F4b). In der Gesamtschau aus quantitativen und qualitativen Daten entsteht jedoch der Eindruck, dass akustische Repräsentationen eher unter die Kategorie *wünschenswert und wichtig* fallen, beziehungsweise – überspitzt formuliert – als *nice to have* bezeichnet werden. Optische Repräsentationen hingegen stellen für das Verstehen aller Schüler*innen eine absolute Notwendigkeit dar.¹²⁸

Hinsichtlich der Frage, inwieweit bei der Bewertung von Erklärqualität Überzeugungen zum Einsatz optischer oder akustischer Repräsentationen eine Rolle spielen (Forschungsfrage F4c), kann anhand der vorgestellten Textstellen eine erste Annahme getroffen werden: Insbesondere dann, wenn eine optische oder akustische Repräsentation fehlt, treten grundsätzliche Überzeugungen der Studierenden wie Lehrenden zu Tage. Diese

128 In weiteren Forschungsvorhaben wäre zu überprüfen, inwieweit diese Beobachtungen auf andere Unterrichtsgegenstände und musikalische Umgangsformen übertragbar sind.

beziehen sich in erster Linie nicht auf einen konkreten methodischen Wunsch bei der Umsetzung, sondern empfehlen grundsätzlich den Einsatz beider Repräsentationsformen. In der Tendenz scheinen die Überzeugungen zum Einsatz optischer Repräsentationsformen vehementer.

Eine Erklärung könnte sein, dass bei der Vermittlung musiktheoretischer Inhalte im Musikunterricht die Verwendung von Notentext und grafischen Darstellungen zur Unterstützung ein gängiges methodisches Mittel darstellt. Während auf inferenzstatistischer Ebene das Fehlen einer akustischen Repräsentation zu keinen signifikanten Unterschieden in der Bewertung der Erklärqualität führt, wird der Einsatz akustischer Repräsentationen in den offenen Antworten dennoch im Verhältnis häufiger erwähnt, als es aufgrund der Globalurteile zu erwarten wäre.

Die getroffene Einschätzung muss vor dem Hintergrund eingeschränkt werden, dass in den untersuchten Erklärungen Hör- oder Klangbeispiele nur gehört werden und es keine Situation gibt, in der Schüler*innen aktiv singen oder musizieren. Dass dieser vor allem forschungsmethodisch begründete Verzicht eine der wenigen Vorgaben an die erklärende Lehrkraft war, erhielten alle Studienteilnehmer*innen zu Beginn der Erhebung als Hinweis. Dennoch gab es einige Äußerungen, in denen die fehlende musikalische Selbsterfahrung der Schüler*innen erwähnt (z. B. *Hsl.3_TA.0a_Bg*; *Hsl.6_NW.o_Bg*) oder Möglichkeiten zur einer solchen gefordert wurden (z. B. *Hsl.11_NW.0a_Vb*; *Stud.28_TA.a_Bg*; *Slk.3_TA.a_Bg*).

Neben eher allgemein formulierten Wünschen nach „handlungs- und schülerorientierten Zugängen“ (*Prof.6_TL.a_Vb*), wie Singen oder Mitspielen, wurde beispielsweise als mögliche Verbesserung vorgeschlagen, „unbedingt über den Körper“ zu gehen und mit „Bewegung“ der Schüler*innen zu arbeiten (*Lk.18_NW.0a_Vb*). In weiteren Vorschlägen sollten Schüler*innen zum Mitklopfen oder Mitzählen angeregt werden (z. B. *Stud.62_TA.a_Vb*; *Prof.4_TA.a_Vb*) oder die Tonleiter durch Vor- und Nachsingen erarbeiten (*Hsl.14_TL.a_Vb*). Generell wurden Wünsche nach gemeinsamen Musizierphasen und dem Sammeln musikalischer Erfahrungen insbesondere an der Stelle des Fragebogens geäußert, an der die Studienteilnehmer*innen Verbesserungsvorschläge unterbreiten sollten (z. B. *Hsl.6_TL.o_Vb*; *Slk.3_NW.0a_Vb*).

5.3.4 Zusammenfassung

Abschließend werden die Antworten zu den drei in Kapitel 5.3 bearbeiteten Forschungsfragen (F2, F3 und F4) zusammengefasst. Mit Blick auf Forschungsfrage zwei, wie sich die Bewertungen von Erklärqualität unter Berücksichtigung der beiden Hauptfaktoren unterscheiden, zeigen sich unabhängig vom jeweiligen Erklärthema signifikante Unterschiede, die einerseits auf die Statusgruppen und andererseits auf unterschiedliche Erklärmodi zurückgeführt werden können. Darüber hinaus liegen für die Globalurteile der Erklärthemen Taktarten und Tonleitern Interaktionseffekte zwischen beiden Faktoren vor. Während dabei die unterschiedlichen Erklärmodi unabhängig der Statusgruppe einen Teil der Varianz aufklären können, sind Unterschiede zwischen den Statusgruppen durch die Erklärmodi bedingt.

Für die Antwort auf Forschungsfrage drei, wie sich die Bewertungen der Erklärqualität zwischen Schüler*innen, Studierenden und Lehrenden unterscheiden, bedeutet dies, dass die Abhängigkeit vom Faktor Erklärqualität zu berücksichtigen ist. Erwartet wurde, dass sich die Bewertungen zwischen Studierenden und Lehrenden aufgrund des Expertisegrades signifikant unterscheiden. Anhand der Ergebnisse der *Post-hoc*-Analysen kann dieser Hypothese zugestimmt werden, da sich die Urteile beider Gruppen in sechs von sieben Erklärungen signifikant unterscheiden. Allerdings zeigten sich bei der grafischen Analyse, dass zwischen allen Videos tendenziell ähnliche Abstände in der Benotung bestehen. Das bedeutet, dass zwar Unterschiede in den absoluten Noten und hinsichtlich des verwendeten Abschnitts auf der Notenskala vorliegen, jedoch die verschiedenen Noten zu allen Erklärungen eines jeden Themas ähnliche Abstände aufweisen. Es kann daher von einem tendenziell ähnlichen Bewertungsmuster beider Gruppen ausgegangen werden, das jedoch von einem unterschiedlichen Referenzwert ausgeht, denn die Lehrenden bewerten alle Erklärungen schlechter als die Studierenden.

Deutlicher unterscheiden sich dazu die Bewertungen der Schüler*innen, die alle Erklärungen auf einem insgesamt kleineren Range bewerten und weniger deutlich zwischen den einzelnen Erklärungen zu einem Thema differieren. Auffällig ist auch, dass nur die Schüler*innen die beiden besten Noten zum Thema Notenwerte vergeben. Da dieses Thema augenscheinlich das am leichtesten zu verstehende und zudem bereits am ehesten schon bekannte der drei Erklärthemen ist, wird davon ausgegangen, dass das Urteil der Schüler*innen von ihrem Vorwissen und der eigenen Einschätzung über das Verstehen beeinflusst ist. Eine weitere Analyse dazu erfolgt in Kapitel 5.5.

Für die Beantwortung von Forschungsfrage vier, wie sich Bewertungen der Erklärqualität in Abhängigkeit vom Erklärmodus unterscheiden, werden zwei verschiedene Datentypen verwendet. Zum einen werden unabhängig vom Thema die Erklärungen, die auf eine optische Repräsentation verzichten, signifikant schlechter beurteilt als solche, in denen diese zum Einsatz kommen. Mit Ausnahme der Bewertungen der Schüler*innen zum Thema Taktarten, treten signifikante Unterschiede in allen Statusgruppen und für die Themen Taktarten und Tonleiter auf. Zum anderen zeigen sich diese Tendenzen auch in den offenen Antworten. Auf Basis einer inhaltsanalytischen Auswertung wurden folgende Beobachtungen gemacht: Zur Begründung der Globalnoten werden akustische und optische Repräsentationen insbesondere dann herangezogen, wenn die jeweilige Repräsentationsform in der Erklärung nicht zum Einsatz kommt. Dabei zeigt sich tendenziell, dass der Einsatz optischer Repräsentationen nicht nur häufiger, sondern auch in größerer Intensität gefordert wird. Empfehlungen zum Einsatz akustischer Repräsentationen liegen in größerer Form vor als es die geringen Unterschiede in den Globalbewertungen erwarten lassen. Neben der Einschränkung, dass singende beziehungsweise musizierende Aktivitäten in den Erklärungen nicht vorgekommen sind – dies wurde insbesondere als mögliche Verbesserung der Erklärqualität vermehrt vorgeschlagen – erscheinen akustische Repräsentationen, die hörend wahrgenommen werden können, eher als wünschenswert oder als *nice to have*; auf sie kann im Zweifel jedoch eher verzichtet werden als auf eine optische Form der Darstellung.

5.4 Qualitätsmerkmale unterrichtlichen Erklärens im Fach Musik

Dieser Abschnitt richtet den Fokus auf Forschungsfrage fünf, welche Merkmale für die Qualität unterrichtlichen Erklärens im Fach Musik ausschlaggebend sind. Dazu werden für alle drei Statusgruppen die Zusammenhänge zwischen den Beurteilungen der ausgewählten Merkmale guten Erklärens sowie der Kontrollvariable *Persönlichkeitswirkung* (als unabhängige Variable) und den Bewertungen zur Erklärqualität (als abhängige Variable) untersucht (siehe Kap. 5.4.1). Auch soll damit die Frage nach einer statusgruppenspezifischen Beurteilung zu Qualitätsmerkmalen beantwortet werden. Nach Angabe der bivariaten Korrelationen zwischen den Merkmalen und der Globalbewertung pro Gruppe erfolgt die Berechnung der drei statusgruppenspezifischen Regressionsanalysen. Anschließend werden Textstellen von Studierenden und Lehrenden ausgewertet, die als offene Antworten vorliegen. Die Ergebnisse der qualitativen Befunde zu den ausgewählten und als deduktive Kategorien vorliegenden Qualitätsmerkmalen sollen einerseits die Befunde der quantitativen Analysen ergänzen (siehe Kap. 5.4.2). Andererseits sollen die offenen Antworten auch weitere potenzielle und in der theoretischen Konzeptualisierung nicht berücksichtigte Qualitätsmerkmale identifizieren (siehe Kap. 5.4.3).¹²⁹

5.4.1 Zusammenhang zwischen Erklärqualität und ausgewählten Merkmalen

F5) Wie hängt die Bewertung der Erklärqualität mit Beurteilungen zu ausgewählten Qualitätsmerkmalen (*Adressatenorientierung, Strukturiertheit, Sprachliche Verständlichkeit, Sprech- und Körperausdruck* sowie *Einsatz musikbezogener Repräsentationen*) zusammen?

Die Beantwortung dieser Frage erfolgt mittels Regressionsanalysen. Untersucht wird, inwieweit die Beurteilungen der einzelnen Qualitätsmerkmale die Qualität der Erklärung (in Form der Globalnote) erklären können. Um einen Überblick der verwendeten Daten zu erhalten, werden zuvor sowohl die deskriptiven Statistiken zu den durchschnittlichen Skalenwerten der einzelnen Merkmale als auch deren bivariate Korrelationen untereinander, einschließlich der Globalnote, getrennt nach Statusgruppen berichtet. Da alle Skalen sowie das Globalurteil auf einer sechsstufigen Skala eingeschätzt werden, können die Analysen in der originären Metrik erfolgen und alle Werte miteinander verglichen werden. Außerdem gilt: Je niedriger der arithmetische Mittelwert, desto höher die Zustimmung zu einzelnen Merkmalen beziehungsweise desto besser die Bewertung der Erklärqualität.

¹²⁹ Beide inhaltsanalytischen Auswertung geben Antworten auf Forschungsfrage sechs (F6), welche Qualitätsmerkmale unterrichtlichen Erklärens in den subjektiven Theorien der Erklärenden eine Bedeutung haben.

Deskriptive Ergebnisse

Die deskriptiven Auswertungen (Mittelwert, Standardabweichung und Spannweite) der einzelnen Qualitätsmerkmale werden getrennt nach Statusgruppe angegeben. Um einen Vergleich der drei einzelnen Gruppen zu ermöglichen, werden zuvor dieselben Kennwerte über alle Studienteilnehmer*innen berichtet (siehe Tab. 5.28).

Tabelle 5.28 Mittelwerte und Standardabweichungen **über alle Studienteilnehmer*innen** zu den fünf Merkmalen guten Erklärens, zur Persönlichkeitswirkung und zum Globalurteil

Erklärthema Erklärmodus Videokürzel	Notenwerte		Taktarten		Tonleiter			Gesamt
	A: opt. NW_o M (SD) [range]	D: opt./aku. NW_oa M (SD) [range]	C: aku. TA_a M (SD) [range]	F: opt./aku. TA_oa M (SD) [range]	E: opt. TL_o M (SD) [range]	G: aku. TL_a M (SD) [range]	B: opt./aku. TL_oa M (SD) [range]	
Strukturiertheit (3 Items)	1,96 (0,82) [1,00; 6,00]	1,95 (0,86) [1,00; 6,00]	2,12 (0,92) [1,00; 6,00]	2,02 (0,91) [1,00; 6,00]	2,05 (0,83) [1,00; 6,00]	2,27 (1,00) [1,00; 6,00]	1,92 (0,82) [1,00; 6,00]	2,04 (0,88)
Adressatenorientierung (3/5 Items)	2,66 (0,99) [1,00; 6,00]	2,76 (0,95) [1,00; 6,00]	2,92 (1,05) [1,00; 6,00]	2,85 (1,08) [1,00; 6,00]	2,84 (0,97) [1,00; 6,00]	3,24 (1,15) [1,00; 6,00]	2,71 (1,04) [1,00; 6,00]	2,85 (1,03)
Sprachliche Verständlichkeit (3 Items)	2,78 (1,01) [1,00; 5,33]	2,74 (1,02) [1,00; 5,33]	2,82 (1,04) [1,00; 5,33]	2,73 (1,04) [1,00; 6,00]	2,74 (0,95) [1,00; 6,00]	2,93 (1,07) [1,00; 6,00]	2,70 (0,99) [1,00; 5,67]	2,78 (1,02)
Sprech- und Körperausdruck (7 Items)	2,05 (0,67) [1,00; 3,86]	2,63 (0,91) [1,00; 6,00]	1,73 (0,63) [1,00; 3,86]	2,11 (0,93) [1,00; 6,00]	2,29 (0,91) [1,00; 6,00]	2,19 (1,00) [1,00; 6,00]	1,92 (0,67) [1,00; 4,57]	2,13 (0,82)
Persönlichkeitswirkung (3 Items)	2,35 (0,84) [1,00; 5,67]	2,78 (1,19) [1,00; 6,00]	2,03 (0,80) [1,00; 5,67]	2,31 (1,14) [1,00; 6,00]	2,35 (1,13) [1,00; 6,00]	2,35 (1,14) [1,00; 6,00]	2,18 (0,82) [1,00; 6,00]	2,34 (1,01)
Einsatz musikbez. Repräsent. (4 Items)	1,80 (0,70) [1,00; 6,00]	1,77 (0,71) [1,00; 5,50]	1,67 (0,80) [1,00; 6,00]	1,59 (0,75) [1,00; 6,00]	1,84 (0,75) [1,00; 5,25]	2,05 (0,84) [1,00; 6,00]	1,64 (0,84) [1,00; 6,00]	1,77 (0,77)
Globalurteil	1,99 (0,96) [0,70; 5,70]	1,83 (0,88) [0,70; 5,70]	2,68 (1,13) [0,70; 6,30]	2,12 (1,14) [0,70; 6,30]	2,13 (1,10) [0,70; 6,30]	2,91 (1,27) [0,70; 6,30]	1,95 (1,02) [0,70; 6,00]	2,23 (1,07)

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Aspekte: 1 bis 6; Skalierung Globalurteil: 0,7 bis 6,3.

Aus Tabelle 5.28 geht hervor, dass alle Studienteilnehmer*innen im Schnitt den *Einsatz musikbezogener Repräsentationen* als wichtig erachten (1,77) und insbesondere die *Strukturiertheit* der Erklärung (2,04) als auch den *Sprech- und Körperausdruck* der Lehrkraft niedriger (2,13) und demnach besser bewerten als *Sprachliche Verständlichkeit* (2,78) und *Adressatenorientierung* (2,85).

Beim Vergleich der einzelnen Erklärungen zeigen sich insbesondere bei den Merkmalen *Adressatenorientierung*, *Sprech- und Körperausdruck* sowie der Kontrollvariable *Persönlichkeitswirkung* größere Unterschiede in den durchschnittlichen Bewertungen. Dabei ist überraschend, dass die deskriptiv größten Abstände für *Sprech- und Körperausdruck* sowie *Persönlichkeitswirkung* vorliegen, obwohl alle Erklärungen von derselben Person gegeben wurden und für diese Aspekte eine größere personenbezogene Stabilität zu erwarten wäre. Für die unterschiedlichen Werte ist zudem kein systematischer Einfluss von Erklärthema und Erklärmodus zu erkennen. Es wird daher angenommen, dass diese Performanzmerkmale nicht nur von der Person abhängig sind. Welche anderen Aspekte dabei eine Rolle spielen, ist in zukünftigen Forschungsvorhaben zu identifizieren. Die Standardabweichungen der einzelnen Beurteilungen liegen über alle Videos hinweg

zwischen 0,67 (*Sprech- und Körperausdruck*, Video NW_o) und 1,15 (*Adressatenorientierung*, Video TL_o). Der Range der Bewertungsskalen wurde in den meisten Fällen vollständig ausgenutzt, weshalb eine hinreichende Sensitivität der Skalen angenommen wird.

Die deskriptiven Kennwerte der Schüler*innen zeigen in Tabelle 5.29 ähnliche Verhältnisse, wie sie bei der Gesamtstichprobe zu beobachten sind. *Strukturiertheit* und *Einsatz musikbezogener Repräsentationen* erfahren die höchste Zustimmung, *Adressatenorientierung* und *Sprachliche Verständlichkeit* die niedrigste. Auch sind innerhalb der Merkmale bei *Sprech- und Körperausdruck* sowie der Kontrollvariable *Persönlichkeitswirkung* die größten Unterschiede über die verschiedenen Erklärungen zu erkennen. Im Unterschied zur Gesamtstichprobe liegen die einzelnen Werte pro Erklärung für das Merkmal *Adressatenorientierung* näher beieinander. Dies kann ein erstes Indiz dafür sein, dass den Schüler*innen die Beurteilung von Tiefenstrukturen des Erklärens, wie dem Grad an *Adressatenorientierung*, schwerer fällt. Außerdem fällt auf, dass die Standardabweichungen im Vergleich zur Gesamtstichprobe tendenziell höher ausfallen und diese Gruppe in ihrer Zusammensetzung größere Heterogenität aufweist.

Tabelle 5.29 Mittelwerte und Standardabweichungen **der Schüler*innen** zu den fünf Merkmalen guten Erklärens, zur Persönlichkeitswirkung und zum Globalurteil

Erklärthema Erklärmodus Videokürzel	Notenwerte		Taktarten		Tonleiter			Gesamt
	A: opt. NW_o M (SD) [range]	D: opt./aku. NW_oa M (SD) [range]	C: aku. TA_a M (SD) [range]	F: opt./aku. TA_oa M (SD) [range]	E: opt. TL_o M (SD) [range]	G: aku. TL_a M (SD) [range]	B: opt./aku. TL_oa M (SD) [range]	
Merkmal								M (SD)
Strukturiertheit (3 Items)	1,87 (0,90) [1,00; 6,00]	1,90 (0,93) [1,00; 6,00]	1,98 (1,03) [1,00; 6,00]	2,00 (1,06) [1,00; 6,00]	2,02 (0,96) [1,00; 6,00]	2,02 (1,01) [1,00; 6,00]	1,91 (0,98) [1,00; 6,00]	1,96 (0,98)
Adressatenorientierung (3/5 Items)	2,58 (1,08) [1,00; 6,00]	2,73 (0,98) [1,00; 6,00]	2,74 (1,07) [1,00; 6,00]	2,91 (1,20) [1,00; 6,00]	2,81 (1,05) [1,00; 6,00]	2,92 (1,13) [1,00; 6,00]	2,75 (1,18) [1,00; 6,00]	2,78 (1,10)
Sprachliche Verständlichkeit (3 Items)	2,75 (1,12) [1,00; 5,00]	2,74 (1,10) [1,00; 5,33]	2,88 (1,17) [1,00; 5,33]	2,83 (1,18) [1,00; 6,00]	2,72 (1,07) [1,00; 6,00]	2,83 (1,15) [1,00; 6,00]	2,74 (1,15) [1,00; 5,67]	2,78 (1,13)
Sprech- und Körperausdruck (7 Items)	2,03 (0,68) [1,00; 3,86]	2,63 (0,99) [1,00; 6,00]	1,86 (0,75) [1,00; 3,86]	2,21 (1,05) [1,00; 6,00]	2,33 (1,04) [1,00; 6,00]	2,29 (1,10) [1,00; 6,00]	1,92 (0,76) [1,00; 4,57]	2,18 (0,91)
Persönlichkeitswirkung (3 Items)	2,23 (0,94) [1,00; 5,67]	2,62 (1,35) [1,00; 6,00]	2,03 (0,91) [1,00; 5,67]	2,32 (1,27) [1,00; 6,00]	2,28 (1,30) [1,00; 6,00]	2,32 (1,27) [1,00; 6,00]	2,17 (0,95) [1,00; 6,00]	2,28 (1,14)
Einsatz musikbez. Repräsent. (4 Items)	1,91 (0,79) [1,00; 6,00]	1,90 (0,80) [1,00; 5,50]	1,81 (0,90) [1,00; 6,00]	1,73 (0,85) [1,00; 6,00]	1,95 (0,84) [1,00; 5,25]	2,15 (0,92) [1,00; 6,00]	1,89 (0,96) [1,00; 6,00]	1,91 (0,87)
Globalurteil	1,90 (1,01) [0,70; 5,70]	1,65 (0,81) [0,70; 5,70]	2,51 (1,10) [0,70; 6,30]	2,32 (1,21) [0,70; 6,30]	2,15 (1,22) [0,70; 6,30]	2,61 (1,27) [0,70; 6,30]	2,10 (1,07) [0,70; 6,00]	2,18 (1,10)

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Aspekte: 1 bis 6; Skalierung Globalurteil: 0,7 bis 6,3.

Auch die Mittelwerte und Standardabweichungen der Gruppe der Studierenden (siehe Tab. 5.30) zeigen im Vergleich zur Gesamtstichprobe zum Großteil ähnliche Muster bezüglich der Merkmale *Strukturiertheit*, *Adressatenorientierung*, *Sprachliche Verständlichkeit* und *Sprech- und Körperausdruck* sowie der Kontrollvariable *Persönlichkeitswirkung*. Im Unterschied zur Gesamtstichprobe bewerten die Studierenden jedoch die Relevanz des *Einsatzes musikbezogener Repräsentationen* für Schülerinnen und Schüler

besser. Auch besteht ein weiterer Unterschied in den insgesamt deutlich niedrigeren Standardabweichungen der Beurteilungen – speziell für den *Einsatz musikbezogener Repräsentationen*.

Tabelle 5.30 Mittelwerte und Standardabweichungen der Studierenden zu den fünf Merkmalen guten Erklärens, zur Persönlichkeitswirkung und zum Globalurteil

Erklärthema Erklärmodus Videokürzel Merkmal	Notenwerte		Taktarten		Tonleiter		Gesamt <i>M</i> (<i>SD</i>)	
	A: opt. NW_o <i>M</i> (<i>SD</i>) [range]	D: opt./aku. NW_oa <i>M</i> (<i>SD</i>) [range]	C: aku. TA_a <i>M</i> (<i>SD</i>) [range]	F: opt./aku. TA_oa <i>M</i> (<i>SD</i>) [range]	E: opt. TL_o <i>M</i> (<i>SD</i>) [range]	G: aku. TL_a <i>M</i> (<i>SD</i>) [range]		B: opt./aku. TL_oa <i>M</i> (<i>SD</i>) [range]
	Strukturiert- heit (3 Items)	1,99 (0,66) [1,00; 3,75]	1,94 (0,53) [1,00; 3,25]	2,36 (0,73) [1,00; 3,75]	2,02 (0,67) [1,00; 3,75]	1,96 (0,53) [1,00; 3,25]		2,55 (0,77) [1,00; 3,75]
Adressatenori- entierung (3/5 Items)	2,49 (0,69) [1,20; 4,40]	2,51 (0,64) [1,00; 4,00]	3,02 (0,88) [1,20; 5,00]	2,50 (0,76) [1,00; 4,20]	2,50 (0,72) [1,00; 4,40]	3,43 (0,98) [1,20; 5,40]	2,31 (0,63) [1,40; 3,80]	2,68 (0,76)
Sprachliche Ver- ständlichkeit (3 Items)	2,79 (0,93) [1,33; 5,33]	2,48 (0,73) [1,00; 4,67]	2,68 (0,87) [1,33; 4,67]	2,40 (0,71) [1,00; 3,67]	2,58 (0,77) [1,33; 4,33]	2,94 (0,91) [1,00; 5,00]	2,52 (0,74) [1,33; 4,33]	2,63 (0,81)
Sprech- und Körperaus- druck (7 Items)	2,14 (0,59) [1,00; 3,29]	2,53 (0,65) [1,29; 3,71]	1,59 (0,53) [1,00; 2,71]	1,90 (0,65) [1,00; 3,00]	2,30 (0,76) [1,00; 3,86]	1,89 (0,56) [1,00; 3,14]	1,97 (0,57) [1,00; 2,86]	2,05 (0,62)
Persönlich- keitswirkung (3 Items)	2,52 (0,80) [1,33; 5,00]	2,71 (0,59) [1,33; 4,00]	1,99 (0,81) [1,00; 4,00]	2,04 (0,68) [1,00; 3,33]	2,27 (0,71) [1,00; 4,33]	2,21 (0,62) [1,00; 3,33]	2,11 (0,80) [1,00; 5,33]	2,26 (0,72)
Einsatz musik- bez. Repräsen- tation. (4 Items)	1,63 (0,45) [1,00; 2,75]	1,57 (0,47) [1,00; 2,75]	1,51 (0,57) [1,00; 3,50]	1,35 (0,47) [1,00; 3,25]	1,61 (0,40) [1,00; 2,50]	1,87 (0,58) [1,00; 3,25]	1,22 (0,35) [1,00; 2,50]	1,54 (0,47)
Globalurteil	1,82 (0,72) [0,70; 4,30]	1,70 (0,62) [0,70; 3,70]	2,71 (1,05) [1,00; 5,00]	1,57 (0,82) [0,70; 4,00]	1,70 (0,62) [0,70; 3,30]	2,86 (1,01) [0,70; 5,30]	1,45 (0,69) [0,70; 3,30]	1,97 (0,79)

Bem.: *M*: Mittelwert; *SD*: Standardabweichung; *range*: Spannweite; Skalierung Aspekte: 1 bis 6; Skalierung Globalurteil: 0,7 bis 6,3.

Abschließend zeigt Tabelle 5.31 die deskriptiven Kennwerte der Lehrenden. Auch in dieser Gruppe fallen die Beurteilungen der Merkmale tendenziell ähnlich zur Gesamtstichprobe aus: *Einsatz musikbezogener Repräsentationen* ist am besten, *Adressatenorientierung* und *Sprachliche Verständlichkeit* am schlechtesten bewertet, wobei im Gegensatz zu den anderen Stichproben größere Unterschiede zwischen den Durchschnittsnoten dieser Merkmale zu beobachten sind. Innerhalb der Skalen *Sprech- und Körperausdruck* sowie der *Persönlichkeitswirkung* liegen auch hier die größten Unterschiede in den Beurteilungen zwischen den einzelnen Erklärungen vor.

Größere Unterschiede zwischen den einzelnen Erklärungen zeigen sich im Gegensatz zu den Beurteilungen der anderen beiden Statusgruppen innerhalb der Merkmale *Einsatz musikbezogener Repräsentationen* und *Adressatenorientierung* (z. B. zwischen den Videos TL_o und TL_oa; siehe Tab. 5.31). Dies kann als Indikator dafür dienen, dass Lehrende aufgrund ihrer gegenüber Schüler*innen und Studierenden höheren Expertise in den Beurteilungen mehr differenzieren. Die einzelnen Standardabweichungen fallen über alle Merkmale und Erklärungen hinweg niedriger aus als in der Gruppe der Schüler*innen aus.

Tabelle 5.31 Mittelwerte und Standardabweichungen **der Lehrenden** zu den fünf Merkmalen guten Erklärens, zur Persönlichkeitswirkung und zum Globalurteil

Erklärthema Erklärmodus Videokürzel	Notenwerte		Taktarten		E: opt. TL_o M (SD)	Tonleiter G: aku. TL_a M (SD)	B: opt./aku. TL_oa M (SD)	Gesamt M (SD)
	A: opt. NW_o M (SD)	D: opt./aku. NW_oa M (SD)	C: aku. TA_a M (SD)	F: opt./aku. TA_oa M (SD)				
Merkmal	[range]	[range]	[range]	[range]	[range]	[range]	[range]	
Strukturiert- heit (3 Items)	2,16 (0,70) [1,00; 4,00]	2,13 (0,96) [1,00; 6,00]	2,25 (0,68) [1,00; 4,00]	2,05 (0,67) [1,00; 4,25]	2,24 (0,67) [1,25; 3,75]	2,68 (0,97) [1,25; 5,75]	2,03 (0,53) [1,00; 3,25]	2,22 (0,74)
Adressatenori- entierung (3/5 Items)	3,09 (0,89) [1,60; 5,40]	3,13 (1,03) [1,40; 5,40]	3,32 (1,07) [1,40; 5,60]	3,07 (0,95) [1,40; 5,40]	3,29 (0,81) [1,40; 5,20]	3,95 (1,00) [2,00; 6,00]	3,08 (0,79) [1,20; 5,00]	3,28 (0,93)
Sprachliche Verständlich- keit (3 Items)	2,87 (0,73) [1,67; 4,33]	3,03 (0,98) [1,33; 5,33]	2,84 (0,82) [1,67; 5,33]	2,78 (0,83) [1,00; 5,33]	2,99 (0,71) [1,67; 5,00]	3,19 (0,96) [1,33; 6,00]	2,81 (0,68) [1,33; 4,33]	2,93 (0,82)
Sprech- und Körperausdruck (7 Items)	1,97 (0,74) [1,00; 3,57]	2,73 (0,91) [1,00; 4,71]	1,65 (0,40) [1,00; 2,43]	2,00 (0,74) [1,00; 3,57]	2,16 (0,59) [1,00; 3,43]	2,16 (0,99) [1,00; 5,29]	1,88 (0,59) [1,14; 3,43]	2,08 (0,71)
Persönlich- keitswirkung (3 Items)	2,40 (0,65) [1,33; 4,00]	3,32 (0,95) [2,00; 5,67]	2,06 (0,58) [1,00; 3,33]	2,54 (1,05) [1,00; 5,33]	2,65 (0,86) [1,00; 4,33]	2,56 (1,09) [1,00; 5,67]	2,27 (0,57) [1,33; 3,67]	2,54 (0,82)
Einsatz musik- bez. Repräsen- t. (4 Items)	1,64 (0,59) [1,00; 3,00]	1,63 (0,58) [1,00; 3,00]	1,44 (0,60) [1,00; 3,50]	1,47 (0,59) [1,00; 3,25]	1,79 (0,71) [1,00; 3,50]	1,95 (0,81) [1,00; 4,50]	1,41 (0,52) [1,00; 3,00]	1,62 (0,63)
Globalurteil	2,35 (0,97) [1,00; 4,30]	2,39 (1,00) [1,00; 5,30]	3,06 (1,21) [1,00; 6,00]	2,19 (1,07) [0,70; 5,00]	2,51 (1,02) [0,70; 5,00]	3,67 (1,22) [1,70; 6,30]	2,07 (1,02) [1,00; 5,00]	2,61 (1,07)

Bem.: M: Mittelwert; SD: Standardabweichung; range: Spannweite; Skalierung Aspekte: 1 bis 6; Skalierung Globalurteil: 0,7 bis 6,3.

Insgesamt können über alle Merkmale und Erklärungen in der Tendenz ähnliche Beurteilungen der einzelnen Gruppen festgestellt werden. Während die Einschätzungen der Schüler*innen eine größere Beurteilungsvarianz aufzeigen, urteilen Studierende und Lehrende tendenziell homogener. Das größte Spektrum an Differenzierung zeigt die Gruppe der Lehrenden in den Beurteilungen – nicht nur zwischen den Merkmalen, sondern auch innerhalb eines Merkmals über verschiedene Erklärungen hinweg. Aufgrund dieser statusgruppenspezifischen Unterschiede werden auch die bivariaten Korrelationen getrennt nach Statusgruppe berichtet. Sie zeigen Zusammenhänge zwischen einzelnen Merkmalen sowie der Globalnote.

Bivariate Korrelationen pro Statusgruppe

Für einen ersten Überblick gibt Tabelle 5.32 die Interskalenkorrelationen der Gesamtstichprobe zwischen allen Merkmalen unterrichtlichen Erklärens, der Kontrollvariable *Persönlichkeitswirkung* sowie dem Globalurteil über alle Videos an. Sie dient als Ausgangs- und Orientierungspunkt für die Darstellung der Interskalenkorrelationen der einzelnen Statusgruppen, die anschließend ausführlich berichtet werden. Auffällig ist zweierlei: Zum einen liegen zwischen allen Variablen signifikant positive Korrelationen vor, zum anderen korrelieren alle Merkmale etwa gleich stark mit dem Globalurteil.

Tabelle 5.32 Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für alle **Studienteilnehmer*innen**

Teilkonstrukt	Strukturiertheit	Adressatenorientierung	Sprech- und Körperausdruck	Sprachliche Verständlichkeit	Persönlichkeitswirkung	Einsatz musikbez. Repräsentationen
Strukturiertheit						
Adressatenorientierung	0,56**					
Sprech- und Körperausdruck	0,28**	0,20**				
Spr. Verständlichkeit	0,32**	0,21**	0,18**			
Persönlichkeitswirkung	0,28**	0,27**	0,70**	0,17**		
Einsatz musikbez. Repräsentationen	0,50**	0,38**	0,19**	0,13**	0,19**	
Globalurteil	0,28**	0,29**	0,26**	0,23**	0,25**	0,26**

Bem.: Korrelationswerte für alle Studienteilnehmer*innen; *: $p \leq 0,05$; **: $p \leq 0,01$.

Tabelle 5.33 zeigt im Folgenden die Interskalenkorrelationen für die Gruppe der Schüler*innen zwischen allen Merkmalen unterrichtlichen Erklärens, der Kontrollvariable *Persönlichkeitswirkung* sowie dem Globalurteil über alle Videos.

Tabelle 5.33 Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für die Gruppe der **Schüler*innen**

Teilkonstrukt	Strukturiertheit	Adressatenorientierung	Sprech- und Körperausdruck	Sprachliche Verständlichkeit	Persönlichkeitswirkung	Einsatz musikbez. Repräsentationen
Strukturiertheit						
Adressatenorientierung	0,49**					
Sprech- und Körperausdruck	0,27**	0,17**				
Spr. Verständlichkeit	0,19**	0,01	0,12**			
Persönlichkeitswirkung	0,26**	0,21**	0,72**	0,08		
Einsatz musikbez. Repräsentationen	0,60**	0,43**	0,18**	0,09**	0,20**	
Globalurteil	0,23**	0,14**	0,31**	0,21**	0,26**	0,26**

Bem.: Korrelationswerte für die Gruppe der Schüler*innen; *: $p \leq 0,05$; **: $p \leq 0,01$.

Mit wenigen Ausnahmen (zwischen *Sprachliche Verständlichkeit* und *Adressatenorientierung* beziehungsweise *Persönlichkeitswirkung*) liegen zwischen allen übrigen Variablen signifikant positive Korrelationen vor. Dabei korrelieren das Merkmal *Sprech- und Körperausdruck* und die Kontrollvariable *Persönlichkeitswirkung* erwartungsgemäß sehr stark ($r = 0,72$). Beide Variablen beschreiben die Performanz der Lehrkraft. Das Merkmal *Strukturiertheit* korreliert stark mit *Einsatz musikbezogener Repräsentationen* ($r = 0,60$). Dies kann ein Hinweis dafür sein, dass sich der Einsatz von Repräsentationen und die Struktur einer Erklärung gegenseitig bedingen. Das Globalurteil korreliert moderat mit dem Merkmal *Sprech- und Körperausdruck*; im Gegensatz zu den anderen Merkmalen weist dies jedoch den höchsten Korrelationskoeffizienten auf. Die generell niedrigen Korrelationen mit der Skala *Sprachliche Verständlichkeit* könnten an den unzureichenden Reliabilitätswerten liegen.

Bei der Gruppe der Studierenden liegen neben wenigen Ausnahmen (zwischen *Adressatenorientierung* und *Sprech- und Körperausdruck* sowie zwischen *Einsatz musikbezogener Repräsentationen* und *Persönlichkeitswirkung*) ausschließlich signifikant positive Korrelationen vor (siehe Tab. 5.34). Erwartungsgemäß und analog zur Gruppe der Schüler*innen korrelieren das Merkmal *Sprech- und Körperausdruck* sowie die Kontrollvariable *Persönlichkeitswirkung* stark miteinander ($r = 0,65$).

Tabelle 5.34 Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für die Gruppe der **Studierenden**

Teilkonstrukt	Strukturiertheit	Adressatenorientierung	Sprech- und Körperausdruck	Sprachliche Verständlichkeit	Persönlichkeitswirkung	Einsatz musikbez. Repräsentationen
Strukturiertheit						
Adressatenorientierung	0,70**					
Sprech- und Körperausdruck	0,18*	0,12				
Spr. Verständlichkeit	0,63**	0,60**	0,21**			
Persönlichkeitswirkung	0,18*	0,17*	0,65**	0,24**		
Einsatz musikbez. Repräsentationen	0,38**	0,34**	0,11	0,26**	0,11	
Globalurteil	0,46**	0,51**	0,15*	0,23**	0,21**	0,25**

Bem.: Korrelationswerte für die Gruppe der Studierenden, *: $p \leq 0,05$; **: $p \leq 0,01$.

Starke Korrelationen liegen auch zwischen den Merkmalen *Adressatenorientierung*, *Strukturiertheit* und *Sprachliche Verständlichkeit* vor (siehe Tab. 5.34). Im Gegensatz zu der Gruppe der Schüler*innen hängen damit insbesondere die Beurteilungen zu Tiefenmerkmalen unterrichtlichen Erklärens miteinander zusammen. Die stärkste Korrelation

mit dem Globalurteil weist das Merkmal *Adressatenorientierung* auf ($r = 0,51$). Sie fällt, wie auch die Korrelation zwischen Globalurteil und *Strukturiertheit*, deutlich höher aus als bei der Gruppe der Schüler*innen.

Die Interskalenkorrelationen der Lehrenden sind zwischen allen Variablen ausschließlich signifikant positiv (siehe Tab. 5.35). Die höchste Korrelation mit dem Globalurteil liegt wie bei den Studierenden für das Merkmal *Adressatenorientierung* vor ($r = 0,47$). Ein Unterschied zeigt sich in den moderaten bis starken Korrelationen zwischen *Sprachliche Verständlichkeit* und den beiden Performanzvariablen *Sprech- und Körperausdruck* ($r = 0,45$) sowie *Persönlichkeitswirkung* ($r = 0,44$).¹³⁰

Tabelle 5.35 Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für die Gruppe der **Lehrenden**

Teilkonstrukt	Strukturiertheit	Adressatenorientierung	Sprech- und Körperausdruck	Sprachliche Verständlichkeit	Persönlichkeitswirkung	Einsatz musikbez. Repräsentationen
Strukturiertheit						
Adressatenorientierung	0,69**					
Sprech- und Körperausdruck	0,46**	0,38**				
Spr. Verständlichkeit	0,64**	0,62**	0,45**			
Persönlichkeitswirkung	0,42**	0,44**	0,70**	0,44**		
Einsatz musikbez. Repräsentationen	0,38**	0,40**	0,16*	0,25**	0,25**	
Globalurteil	0,32**	0,47**	0,21**	0,26**	0,17*	0,29**

Bem.: Korrelationswerte für die Gruppe der Lehrenden; *: $p \leq 0,05$; **: $p \leq 0,01$.

Zusammenfassend liefern die bivariaten Interskalenkorrelationen erste deskriptive Hinweise auf potenzielle Zusammenhänge zwischen einzelnen Merkmalen, der Kontrollvariable sowie dem Globalurteil pro Statusgruppe. Erstens korrelieren in allen drei Gruppen *Sprech- und Körperausdruck* sowie *Persönlichkeitswirkung* erwartungsgemäß stark miteinander. In den nachfolgenden regressionsanalytischen Verfahren ist daher von Interesse, inwieweit sich beide Variablen Varianz hinsichtlich der Vorhersage der Globalnote teilen.

Zweitens korrelieren die Merkmale innerhalb der einzelnen Gruppen unterschiedlich stark miteinander: Während *Strukturiertheit* bei den Schüler*innen stark mit *Einsatz musikbezogener Repräsentationen* sowie *Adressatenorientierung* zusammenhängt (siehe

¹³⁰ An anderer Stelle (siehe Frei et al., i. Vorb.) werden Interskalenkorrelationen für die Schüler*innen und die Gruppe der Erklärenden – Studierende und Lehrende zusammengefasst – berichtet. Dies erfolgt dort, da entsprechend den beiden Perspektiven auf unterrichtliches Erklären zwei Regressionsanalysen durchgeführt wurden.

Tab. 5.33), korrelieren bei Studierenden und Lehrenden die drei tiefenstrukturellen Merkmale, *Strukturiertheit*, *Adressatenorientierung* und *Sprachliche Verständlichkeit*, hoch miteinander (siehe Tab. 5.34 und Tab. 5.35). Dies kann als Indiz für Unterschiede in der Bedeutung einzelner Merkmale zwischen den Gruppen gewertet werden und wird anhand der Regressionsanalysen inferenzstatistisch untersucht.

Drittens zeigen sich Unterschiede insofern, als bei der Gruppe der Schüler*innen insbesondere die Merkmale der Oberflächenebene wie *Sprech- und Körperausdruck*, *Persönlichkeitswirkung* und *Einsatz musikbezogener Repräsentationen* moderat mit der Globalnote zusammenhängen (siehe Tab. 5.33). Bei Studierenden und Lehrenden hingegen korrelieren vor allem *Adressatenorientierung* und *Strukturiertheit* hoch mit der Globalnote (siehe Tab. 5.34 und Tab. 5.35). Anhand der folgenden Regressionsanalysen soll daher eine Aussage getroffen werden, inwieweit für die Vorhersage des Globalurteils zwischen den einzelnen Gruppen relevant ist, ob die Prädiktoren eher Oberflächen- oder Tiefenstrukturen zuzuordnen sind.

Linear Mixed Models

Mithilfe hierarchisch linearer Regressionsmodelle (siehe Hilbert et al., 2019) soll nun abgeschätzt werden, welche Bedeutung die ausgewählten Merkmale unterrichtlichen Erklärens im Fach Musik für die Bewertung der Erklärqualität haben. Unter wechselseitiger Kontrolle der einzelnen Merkmale kann somit untersucht werden, inwieweit die unabhängigen Variablen als Prädiktoren die Globalurteile zur Erklärqualität als abhängiges Kriterium vorhersagen können.

Damit werden Antworten auf Forschungsfrage fünf gegeben, die nach Zusammenhängen zwischen den Qualitätsmerkmalen und der Erklärqualität fragt (siehe Kap. 3.2.3). Von Interesse sind dabei auch Gemeinsamkeiten und Unterschiede zwischen den drei Statusgruppen, weshalb für Schüler*innen, Studierende und Lehrende jeweils einzelne Modelle berichtet werden. Dabei wird erwartet, dass *Adressatenorientierung* ein zentrales Merkmal unterrichtlichen Erklärens für Schüler*innen (*Hypothese 5.1*), Studierende (*Hypothese 5.2*) und Lehrende (*Hypothese 5.3*) darstellt und signifikante Zusammenhänge jeweils zwischen den Merkmalsbeurteilungen und der Bewertung der Erklärqualität vorliegen.¹³¹ Die Daten sind für jedes Modell nach Personen und Videos geschachtelt, um deren personen- und videobezogene Struktur zu berücksichtigen.

¹³¹ Um die Ergebnisse der Mehrebenenanalysen einordnen zu können, werden hier die einzelnen verwendeten Kennwerte erklärt. Die Intraklassenkorrelation (*ICC*; Field, 2013, S. 729) dient als Maß dafür, welchen Anteil an der Varianz im Schnitt anhand der Personen- und Videostruktur erklärt werden kann. Anhand dessen kann beispielsweise berücksichtigt werden, dass einzelne Personen grundsätzlich bessere oder schlechtere Bewertungen für einzelne Videos abgeben oder einzelne Videos generell besser oder schlechter bewertet werden. Die Regressionskoeffizienten b sind unstandardisiert und können in ihrer originalen Metrik interpretiert werden, da sowohl die abhängige Variable Globalurteil als auch die unabhängigen Variablen, alle Merkmale sowie die Kontrollvariable, auf einer sechsstufigen Skala bewertet wurden. So würde der Regressionskoeffizient $b = 0,10$ eines Merkmals bedeuten, dass sich die Globalbeurteilung im Schnitt um die Einheit 0,10 verbessert, wenn das Merkmal um durchschnittlich eine Einheit besser bewertet wird. Die Determinationskoeffizienten geben den Anteil der durch die Prädiktoren aufgeklärten Varianz des Globalurteils an. Dabei

Schüler*innen

Tabelle 5.36 zeigt die Ergebnisse des hierarchisch linearen Regressionsmodells für die Gruppe der Schüler*innen mit der abhängigen Variable Globalurteil und den unabhängigen Variablen, bestehend aus den Qualitätsmerkmalen sowie der *Persönlichkeitswirkung* als Kontrollvariable. Bei drei Prädiktoren können in diesem Modell¹³² signifikante Zusammenhänge mit dem Kriterium erkannt werden: Den größten signifikanten Einfluss hat das Merkmal *Sprech- und Körperausdruck* mit einem Regressionsgewicht von 0,25. Da in allen sieben Videos dieselbe Lehrkraft erklärt und für dieses personale Merkmal eine höhere Stabilität in der Beurteilung naheliegender wäre, ist die hohe Bedeutung gegenüber anderen Qualitätsmerkmalen überraschend.

Tabelle 5.36 Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach **Schüler*innen** ($N_{max} = 148$) und den sieben Videos geordneten Datenstruktur

Schüler*innen ($N_{Beob.} = 480$) ¹³³		
ICC	%	39,52
Konstante	<i>b</i>	0,58*
Strukturiertheit	<i>b</i>	0,09
Adressatenorientierung	<i>b</i>	0,01
Sprech- und Körperausdruck	<i>b</i>	0,25**
Sprachliche Verständlichkeit	<i>b</i>	0,12**
Persönlichkeitswirkung	<i>b</i>	0,05
Einsatz musikbez. Repräsentationen	<i>b</i>	0,18**
R^2 (marginal konditional)		0,16 0,39

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen ($N \times 7$ Videos); ICC: Intraklassenkorrelation; *b*: unstandardisierter Regressionskoeffizient; R^2 : Determinationskoeffizient; *: $p \leq 0,05$; **: $p \leq 0,01$.

Keinen signifikanten Einfluss auf Unterschiede im Globalurteil nimmt die Beurteilung der Kontrollvariable *Persönlichkeitswirkung* trotz einer hohen Korrelation mit *Sprech- und Körperausdruck* ein (siehe Tab. 5.33). Der *Einsatz musikbezogener Repräsentationen* hat für Schüler*innen ebenso eine signifikante Bedeutung ($b = 0,18$) hinsichtlich der Bewertung von Erklärqualität wie das Merkmal *Sprachliche Verständlichkeit*, wenngleich dieses mit einem Regressionsgewicht von 0,12 im Vergleich schwächer ausgeprägt

bezieht sich der Wert R^2 *marginal* auf den Einfluss aller festen Effekte, also den Prädiktoren, die in das Modell eingehen. R^2 *konditional* gibt an, welcher Anteil der Gesamtvarianz anhand dieser festen sowie zusätzlich auch der variablen Effekte erklärt werden kann, die auf systematische personenbezogene und videobezogene Varianz zurückzuführen ist.

132 Die Intraklassenkorrelation beträgt 39,52 %.

133 Die Anzahl an Beobachtungen ist reduziert, da für das Merkmal *Sprech- und Körperausdruck* von jeder Person nur zu drei beziehungsweise vier Videos Einschätzungen vorliegen. Demnach kann nur auf Basis dieser Zahlen die Regressionsanalyse durchgeführt werden.

ist. Zu berücksichtigen sind bei der Interpretation jedoch die niedrigen Reliabilitätswerte der Skala (siehe Tab. 5.3), weshalb die Aussagekraft für dieses Merkmal limitiert ist.¹³⁴

Ohne Berücksichtigung Sprachlicher Verständlichkeit liegen mit Einsatz musikbezogener Repräsentationen sowie Sprech- und Körperausdruck zwei Merkmale vor, die auf einer Oberflächenebene verortet sind. Es liegt die Vermutung nahe, dass Schüler*innen diese leichter beantworten können und eher das Verhalten der Lehrkraft als die Inhalte der Erklärung bewerten.

Für die beiden Merkmale *Strukturiertheit* und *Adressatenorientierung* kann kein signifikanter Einfluss auf die Bewertung der Erklärqualität für die Schüler*innen nachgewiesen werden, für die Beurteilung des Merkmals *Adressatenorientierung* liegt sogar nahezu eine Nullkorrelation mit dem Globalurteil vor. Demnach kann *Hypothese 5.1*, dass insbesondere *Adressatenorientierung* für die Schüler*innen einen signifikanten Einfluss auf deren Bewertung der Erklärqualität hat, nicht unterstützt werden. Nachdem jedoch die Merkmale *Strukturiertheit* und *Einsatz musikbezogener Repräsentationen* hoch korreliert sind (siehe Tab. 5.33), kann vermutet werden, dass die fehlende Signifikanz von *Strukturiertheit* unter anderem auf die geteilte Varianz zwischen den beiden Prädiktoren zurückzuführen ist. Auch können die fehlenden signifikanten Effekte darin begründet sein, dass Schüler*innen mit der Beurteilung von Merkmalen auf der Tiefenebene Probleme haben, da die dazugehörigen Items (siehe Tab. 4.7 für *Strukturiertheit*, Tab. 4.8 für *Adressatenorientierung*) hauptsächlich als hoch-inferent zu klassifizieren und damit schwerer zu beurteilen sind. In der Gesamtschau aller Beurteilungen von Qualitätsmerkmalen erscheint diese Vermutung sowohl für Oberflächen- wie Tiefenmerkmale unterrichtlichen Erklärens plausibel. Die Varianzaufklärung dieses linearen Modells beträgt 16 % für die festen Effekte und 39 % für die festen gemeinsam mit den zufälligen Effekten (siehe Tab. 5.36), was als zufriedenstellend einzuschätzen ist.

Studierende

Für die Gruppe der Studierenden werden die Ergebnisse zum Mehrebenenmodell¹³⁵ in Tabelle 5.37 berichtet. Im Vergleich zu den Schüler*innen zeigen sich bei den Studierenden andere signifikante Prädiktoren für die Erklärqualität. Die größte Bedeutung kann dem Merkmal *Strukturiertheit* beigemessen werden. Es liegt ein signifikanter Effekt mit einem Regressionskoeffizienten von 0,38 vor. Außerdem besitzt das Merkmal *Adressatenorientierung* einen signifikanten Effekt, der das Globalurteil mit einem Regressionskoeffizienten von 0,32 vorhersagen kann. Damit kann *Hypothese 5.2*, dass *Adressatenorientierung* für Studierende ein relevantes Merkmal von Erklärqualität darstellt, unterstützt werden.

134 Neben der geringen Anzahl an Items zeigen die *part-whole*-korrigierten Itemtrennschärfen bei den Schüler*innen, dass für jedes Video lediglich zwei Items zu einer Skala zusammengefasst werden können und diese überdies einen formativen Charakter aufweisen.

135 Bei den Studierenden beträgt die Intraklassenkorrelation 24,70 %.

Tabelle 5.37 Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach **Studierenden** ($N_{max} = 61$) und den sieben Videos geordneten Datenstruktur

Studierende ($N_{Beob.} = 186$)		
ICC	%	24,70
Konstante	<i>b</i>	0,60*
Strukturiertheit	<i>b</i>	0,38**
Adressatenorientierung	<i>b</i>	0,32**
Sprech- und Körperausdruck	<i>b</i>	0,07
Sprachliche Verständlichkeit	<i>b</i>	-0,20*
Persönlichkeitswirkung	<i>b</i>	0,06
Einsatz musikbez. Repräsentationen	<i>b</i>	-0,06
R^2 (marginal konditional)		0,22 0,35

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen ($N \times 7$ Videos); ICC: Intraklassenkorrelation; *b*: unstandardisierter Regressionskoeffizient; R^2 : Determinationskoeffizient; *: $p \leq 0,05$; **: $p \leq 0,01$.

Auffällig ist, dass gerade die beiden eben genannten Merkmale die Globalbewertung der Studierenden signifikant beeinflussen, die im linearen Modell der Schüler*innen keine (statistisch) signifikante Bedeutung besitzen. Im umgekehrten Fall zeigt sich bei der Gruppe der Studierenden für die Merkmale, die eher auf einer Oberflächenebene zu verorten sind, kein statistisch signifikanter Einfluss: Weder können *Sprech- und Körperausdruck* noch *Einsatz musikbezogener Repräsentationen* Unterschiede in der Globalbeurteilung erklären.

Auch der Kontrollvariable *Persönlichkeitswirkung* wird von den Studierenden keine (statistisch) signifikante Bedeutung beigemessen. Das bedeutet, dass für die Studierenden die erklärende Person weniger im Fokus steht als inhaltliche und strukturelle Aspekte der Erklärung. Da der *Einsatz musikbezogener Repräsentationen* moderate Korrelationen zu beiden Merkmalen *Strukturiertheit* und *Adressatenorientierung* aufweist (siehe Tab. 5.34), kann geteilte Varianz vermutet werden. Das bedeutet, dass der *Einsatz musikbezogener Repräsentationen* durchaus relevant sein kann, sich dies aufgrund von Zusammenhängen mit den Beurteilungen zu *Strukturiertheit* und *Adressatenorientierung* im Regressionsmodell nicht zeigt.

Ein überraschendes und bemerkenswertes Ergebnis zeigt sich für den Einfluss des Merkmals *Sprachliche Verständlichkeit*. Im Gegensatz zur Gruppe der Schüler*innen liegen für die Skala bei den Studierenden zufriedenstellende Reliabilitätswerte vor, sodass die Interpretation des Regressionskoeffizienten im Mehrebenenmodell keiner Limitation unterliegt. Ein negatives Regressionsgewicht von $-0,20$ bedeutet in diesem Fall, dass die Erklärqualität umso besser bewertet wurde, je geringer die Zustimmung für die Skala *Sprachliche Verständlichkeit* ausfällt.¹³⁶ Dieser Befund ist kontraintuitiv, denn er

¹³⁶ Dieser Befund zeigt sich auch im Regressionsmodell für die Studierenden in der FALKE-q-D-Studie im Fach Deutsch (Gaier, 2020, S. 209).

bedeutet mit Blick auf die einzelnen Items der Skala, dass die Verwendung unbekannter Fachbegriffe, die Formulierung tendenziell zu langer und damit komplexerer Sätze und eine eher niedrige *Sprachliche Verständlichkeit* zu einer höheren Erklärqualität führen.¹³⁷ Als ein möglicher Grund wäre denkbar, dass eine komplexere Sprache für die Studierenden einen fachlich-intellektuell höheren Gehalt der Erklärung suggeriert. Dieser wird mit einer steigenden Erklärqualität in Verbindung gebracht, ohne dass dabei die Auswirkungen auf das Verstehenspotenzial für Schüler*innen berücksichtigt und fachdidaktisch relevante Aspekte wie eine für Lernende leicht verständliche Sprache überstrahlt werden.¹³⁸

Die Varianzaufklärung von 22 % für die festen Effekte und 35 % für die festen gemeinsam mit den zufälligen Effekten, die im Modell der Studierenden erzielt wird, kann als gut eingestuft werden. Dennoch sind für die Gruppe der Studierenden weitere Merkmale für die Bewertung der Erklärqualität relevant. Erste Hinweise dazu kann die inhaltsanalytische Auswertung der offenen Antworten liefern (siehe Kap. 5.4.3).

Lehrende

Tabelle 5.38 zeigt die Ergebnisse des hierarchisch linearen Regressionsmodells¹³⁹ für die Gruppe der Lehrenden. Im Gegensatz zu den beiden Modellen der Schüler*innen und der Studierenden kann für die Gruppe der Lehrenden nur das Merkmal *Adressatenorientierung* Unterschiede in den Globalurteilen signifikant erklären. Dabei liegt mit 0,48 der höchste Regressionskoeffizient aller drei Modelle vor. Somit kann *Hypothese 5.3* unterstützt werden, dass Adressatenorientierung ein bedeutendes Qualitätsmerkmal unterrichtlichen Erklärens für die Gruppe der Lehrenden darstellt. Es liegt die Vermutung nahe, dass die Orientierung an Kenntnissen und Fähigkeiten der Schüler*innen von den Lehrenden als Leitziel für unterrichtliches Erklären verstanden wird, das weitere Merkmale subsummiert. Diese Annahme kann anhand der durchweg mittleren bis hohen Korrelationen zwischen *Adressatenorientierung* und den weiteren Merkmalen gestützt werden (siehe Tab. 5.35). Zudem würde das als erster Indikator für die Validierung des Modells guten Erklärens im Musikunterricht dienen (siehe Abb. 2.2, Kap. 2.4.2; siehe auch Frei et al., 2022), welches *Adressatenorientierung* als Leitziel unterrichtlichen Erklärens versteht und auf Basis einer fachübergreifenden Perspektive theoretisch konzeptualisiert wurde (siehe Kap. 2.3.2).

137 Es wurde überprüft, ob ein Suppressionseffekt für das Kriterium der *Sprachlichen Verständlichkeit* vorliegt (Eid et al., 2017; siehe Valicer, 1978). Dabei konnte kein Effekt nachgewiesen werden.

138 Eine sprachwissenschaftliche Analyse, die an die einzelnen Items der Skala geknüpft ist, wird für alle im Projekt FALKE-q beteiligten Fächer, bei Thim-Mabrey und Lindl (i. Vorb.) vorgenommen.

139 Die Intraklassenkorrelation liegt für das Modell der Lehrenden bei 50,83 %. Dieser hohe Wert bedeutet, dass die Hälfte der Merkmalsvarianz auf die Personen- und Videostruktur zurückzuführen ist. Im Vergleich zu den Schüler*innen und Studierenden können demnach stabilere Bewertungsmaßstäbe für die Stichprobe der Lehrenden hinsichtlich der ausgewählten Erklärungen vermutet werden.

Tabelle 5.38 Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach **Lehrenden** ($N_{max} = 62$) und den sieben Videos geordneten Datenstruktur

Lehrende ($N_{Beob.} = 169$)		
<i>ICC</i>	%	50,83
Konstante	<i>b</i>	0,61
Strukturiertheit	<i>b</i>	0,06
Adressatenorientierung	<i>b</i>	0,48**
Sprech- und Körperausdruck	<i>b</i>	0,01
Sprachliche Verständlichkeit	<i>b</i>	0,01
Persönlichkeitswirkung	<i>b</i>	-0,03
Einsatz musikbez. Repräsentationen	<i>b</i>	0,15
<i>R</i> ² (marginal konditional)		0,25 0,39

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen ($N \times 7$ Videos); *ICC*: Intraklassenkorrelation; *b*: unstandardisierter Regressionskoeffizient; R^2 : Determinationskoeffizient; *: $p \leq 0,05$; **: $p \leq 0,01$.

Für die Merkmale *Strukturiertheit*, *Sprech- und Körperausdruck*, *Sprachliche Verständlichkeit* und *Persönlichkeitswirkung* sind die Regressionsgewichte alle nahezu nullwertig. Das verwundert insbesondere für die Merkmale *Strukturiertheit* und *Sprachliche Verständlichkeit*, da beide mit *Adressatenorientierung* hoch korrelieren und daher von einem hohen Anteil geteilter Varianz zwischen den Merkmalen auszugehen ist (siehe Tab. 5.35). Der *Einsatz musikbezogener Repräsentationen* hat einen etwas höheren, jedoch ebenfalls statistisch nicht signifikanten Einfluss mit einem Regressionskoeffizient von $b = 0,15$.

Die Varianzaufklärung liegt für dieses Modell bei 25 % für die festen Effekte und 39 % für die festen gemeinsam mit den zufälligen Effekten. Beide Werte können als gut bezeichnet werden, wobei auch weitere bisher unberücksichtigte Merkmale die Qualität unterrichtlichen Erklärens im Fach Musik aus Sicht der Lehrenden beeinflussen. Wie bei den Studierenden werden diesbezüglich weitere Erkenntnisse durch eine inhaltliche Analyse der offenen Antworten erwartet (siehe Kap. 4.5.3).

Vergleich der drei Regressionsmodelle

Ein Vergleich der nach Statusgruppen getrennten hierarchisch linearen Regressionsmodelle (siehe Tab. 5.39)¹⁴⁰ fasst die zentralen Ergebnisse zusammen:

Erstens können die aufgestellten Hypothesen zur Relevanz des Merkmals *Adressatenorientierung* nur für die Gruppen der Studierenden (*Hypothese 5.2*) und Lehrenden (*Hypothese 5.3*) aufrechterhalten werden. Zweitens zeigen einzelne Merkmale nur in jeweils einer Statusgruppe einen statistisch bedeutsamen Einfluss auf das Globalurteil: Während im Modell der Studierenden das Merkmal *Strukturiertheit* einen signifikanten Einfluss aufweist, sind für die Schüler*innen insbesondere *Sprech- und Körperausdruck*

¹⁴⁰ Die Zusammenfassung der drei Modelle aus den Tabellen 5.36, 5.37 und 5.38 dient der besseren Übersichtlichkeit.

sowie der *Einsatz musikbezogener Repräsentationen* relevant. Drittens ist vor diesem Hintergrund bemerkenswert, dass keines der Qualitätsmerkmale von allen drei Gruppen als relevant erachtet wird, weshalb von einer perspektivenspezifischen Validität der Qualitätsmerkmale unterrichtlichen Erklärens ausgegangen werden muss. Viertens können die Ergebnisse zur *Sprachliche Verständlichkeit* nur mit Vorsicht interpretiert werden. Zum einen limitieren niedrige Reliabilitätswerte für die Skala der Schüler*innen die Aussagekraft. Zum anderen liegen bei den Studierenden kontraintuitive Befunde vor. Hier bleibt unklar, ob die Beurteilungen zu *Sprachliche Verständlichkeit* einer Erklärung nicht von anderen Faktoren überdeckt wurden. Bei den Lehrenden hingegen spielt sie mit Blick auf die Erklärqualität keine Rolle. Fünftens zeigt die Kontrollvariable *Persönlichkeitswirkung* in allen drei Modellen keinen signifikanten Einfluss auf die Globalnote.

Tabelle 5.39 Übersicht der drei hierarchischen linearen Regressionsmodelle getrennt nach Statusgruppe

		Schüler*innen ($N_{Beob.} = 480$)	Studierende ($N_{Beob.} = 186$)	Lehrende ($N_{Beob.} = 169$)
<i>ICC</i>	%	39,52	24,70	50,83
Konstante	<i>b</i>	0,58*	0,60*	0,61
Strukturiertheit	<i>b</i>	0,09	0,38**	0,06
Adressatenorientierung	<i>b</i>	0,01	0,32**	0,48**
Sprech- und Körperausdruck	<i>b</i>	0,25**	0,07	0,01
Sprachliche Verständlichkeit	<i>b</i>	0,12**	-0,20*	0,01
Persönlichkeitswirkung	<i>b</i>	0,05	0,06	-0,03
Einsatz musikbez. Repräsentationen	<i>b</i>	0,18**	-0,06	0,15
<i>R</i> ² (marginal konditional)		0,16 0,39	0,22 0,35	0,25 0,39

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen ($N \times 7$ Videos); *ICC*: Intraklassenkorrelation; *b*: unstandardisierter Regressionskoeffizient; *R*²: Determinationskoeffizient; *: $p \leq 0,05$; **: $p \leq 0,01$.

Zusammenfassung

Auf Forschungsfrage fünf (F5) können demnach folgende Antworten gegeben werden: Zwar sind für die verschiedenen Statusgruppen unterschiedliche Merkmale bedeutsam, jedoch zeigen alle der fünf ausgewählten Qualitätsmerkmale unterrichtlichen Erklärens statistisch signifikante Einflüsse auf die jeweiligen Globalurteile zur Erklärqualität. Dies wird auch deutlich, wenn ein lineares Modell¹⁴¹ über die gesamte Stichprobe gerechnet wird (siehe Tab. 5.40). Darin werden alle Merkmale mit Ausnahme der Kontrollvariable *Persönlichkeitswirkung* signifikant. Dies kann als erster Indikator für Konstruktvalidität der Qualitätsmerkmale unterrichtlichen Erklärens gedeutet werden. Da die Varianzaufklärung für die festen Effekte mit 16 % einen zufriedenstellenden, jedoch im Vergleich

141 Die Intraklassenkorrelation liegt bei 39,09 %.

zu den einzelnen Modellen geringeren Wert¹⁴² aufweist, erscheint die Aufteilung nach Statusgruppen für einen höheren Erkenntnisgewinn notwendig. Der Wert für die festen gemeinsam mit den zufälligen Effekten liegt mit 43 % erwartungsgemäß zwischen den einzelnen Werten der drei Modelle.

Tabelle 5.40 Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach **allen Studienteilnehmer*innen** ($N_{max} = 271$) und den sieben Videos geordneten Datenstruktur

Gesamtstichprobe ($N_{Beob.} = 835$)		
<i>ICC</i>	%	39,09
Konstante	<i>b</i>	0,45*
Strukturiertheit	<i>b</i>	0,12*
Adressatenorientierung	<i>b</i>	0,16**
Sprech- und Körperausdruck	<i>b</i>	0,17**
Sprachliche Verständlichkeit	<i>b</i>	0,11*
Persönlichkeitswirkung	<i>b</i>	0,06
Einsatz musikbez. Repräsentationen	<i>b</i>	0,12*
<i>R</i> ² (marginal konditional)		0,16 0,43

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen ($N \times 7$ Videos); *ICC*: Intraklassenkorrelation; *b*: unstandardisierter Regressionskoeffizient; *R*²: Determinationskoeffizient; *: $p \leq 0,05$; **: $p \leq 0,01$.

5.4.2 Subjektive Theorien¹⁴³ zu den ausgewählten Merkmalen

F6) Welche Merkmale unterrichtlichen Erklärens erachten die Studienteilnehmer*innen hinsichtlich der Bewertung von Erklärqualität als relevant?

Das vorliegende Kapitel gibt Antworten auf die Frage, welche Bedeutung die fünf ausgewählten Merkmale *Adressatenorientierung*, *Strukturiertheit*, *Sprachliche Verständlichkeit*, *Sprech- und Körperausdruck* sowie *Einsatz musikbezogener Repräsentationen* für die Qualität unterrichtlichen Erklärens aus Sicht der Studierenden und Lehrenden haben (Forschungsfrage F6a).¹⁴⁴

142 Das Modell der Schüler*innen (siehe Tab. 5.36) zeigt auch einen *R*²-Wert von 0,16.

143 Der Begriff meint hier individuelle Sichtweisen der Studierenden und Lehrenden, die in Form offener Antworten zur Qualität der Erklärungen abgegeben wurden. Von Interesse sind die jeweiligen individuellen Vorstellungen zum guten Erklären unabhängig davon, ob sich diese auf eigene Erfahrungen, (musikdidaktische) Konzeptionen oder andere Quellen stützen. Nicht gemeint ist ein Verständnis des Begriffs, wie er dem „Forschungsprogramm Subjektive Theorien“ (Groeben et al., 1988) zugrunde liegt. Gründe dazu werden bei Niessen (2006a) diskutiert.

144 Nicht berücksichtigt werden an dieser Stelle weitere potenzielle Qualitätsmerkmale, die in den offenen Antworten genannt wurden, jedoch den fünf ausgewählten Kriterien nicht zugeordnet werden können. Diese sind Inhalt von Kapitel 5.4.3 und sollen ein möglichst vollständiges Bild unterrichtlichen Erklärens ergeben, das zudem blinde Flecken der Konzeptualisierung auszuleuchten vermag.

Grundlage der Darstellung sind deren subjektive Theorien zu Qualitätsmerkmalen unterrichtlichen Erklärens. Dazu werden die offenen Antworten aus beiden Statusgruppen inhaltsanalytisch untersucht. Beim Bericht der Ergebnisse wird in der Regel auf eine Unterscheidung der beiden Statusgruppen, außer es wird explizit darauf hingewiesen, verzichtet. Aufgrund des explorativen Ansatzes wird keine Wertung hinsichtlich der Qualität der Aussagen vorgenommen.

Durch den gleichwertigen Einbezug aller Antworten wird das Ziel verfolgt, ein möglichst facettenreiches Bild der einzelnen Qualitätsmerkmale zu entwickeln, das jedoch nicht an jeder Stelle gleichermaßen hochauflösend sein kann. Einbezogen werden zusätzlich alle Aussagen, die sich auf die Qualität unterrichtlichen Erklärens beziehen und an drei Stellen des Fragebogens abgefragt wurden.¹⁴⁵ Einzig für die Validierung der Merkmale unterrichtlichen Erklärens richtet sich der Blick explizit auf die Angaben, mit denen die Bewertung des Globalurteils in Durchgang eins begründet wurde. Da zu diesem Zeitpunkt im Fragebogen noch keine merkmalsbezogenen Items abgefragt wurden, kann anhand der Begründungen eine Einschätzung über die Konstruktvalidität der ausgewählten fünf Merkmale unterrichtlichen Erklärens abgegeben werden. Die Darstellung erfolgt getrennt nach Qualitätsmerkmalen.

Adressatenorientierung

Von den insgesamt 3.374 Codierungen entfallen 268 codierte Stellen auf das Merkmal *Adressatenorientierung*, von denen wiederum 190 als Begründung zum Globalurteil abgegeben wurden. Aufgrund dessen wird angenommen, dass Adressatenorientierung für Studierende und Lehrende ein Qualitätsmerkmal unterrichtlichen Erklärens darstellt. Zudem liegen inhaltliche Überschneidungen zwischen der Operationalisierung der Skala *Adressatenorientierung* und den acht Subkategorien, mit denen die Textstellen codiert wurden, vor (siehe *Kategoriensystem 2b, Anhang F4*). So spielt in den offenen Antworten der Studienteilnehmer*innen¹⁴⁶ insbesondere die Frage, inwieweit in den Erklärungen das Vorwissen der Schüler*innen sowie deren Lebenswelt berücksichtigt werden, eine wichtige Rolle.¹⁴⁷

Der Alltags- und Lebensweltbezug einer Erklärung wird im Kontext des bekannten Fußballrhythmus-Beispiels (Thema Taktarten) und dem geläufigen Konzept der Zeitstunde (Thema Notenwerte) angesprochen. Die Einschätzungen darüber, dass ein bekannter Rhythmus verfremdet wird, fallen heterogen aus: Einerseits wird positiv bewertet,

145 Da allgemeine Bewertungen wie „logischer Aufbau“ (Lk.20_TL.oa_Bg) oder „wenig schülerorientiert“ (Prof.6_TA.a_Bg) zwar für die Gesamtbetrachtung eines Merkmals wichtig, jedoch hinsichtlich einer differenzierten Darstellung einzelner Facetten nicht zielführend sind, werden diese Codierungen in der Auswertung nicht explizit aufgeschlüsselt, fließen jedoch in die Konstruktvalidierung der Qualitätsmerkmale ein.

146 Wenn in diesem sowie im folgenden Kapitel von Studienteilnehmer*innen die Rede ist, sind stets nur Studierende und Lehrende gemeint. Von der Gruppe der Schüler*innen liegen keine offenen Antworten vor.

147 Der dritte Aspekt, inwieweit die Erklärungen didaktisch reduziert waren, wurde hingegen kaum thematisiert und wird daher nicht in die Auswertung einbezogen.

dass das gewählte Fußball-Rhythmusbeispiel bei den Erklärungen zu Taktarten bekannt ist (z. B. *Slk.4_TA.aa_Bg*; *Prof.8_TA.a_Bg*; *Lk.16_TA.a_Bg*) und dadurch sowohl die Anschaulichkeit der Erklärung steigert (*Lk.9_TA.aa_Bg*) als auch das Interesse der Schüler*innen weckt (*Hsl.5_TA.aa_Bg*). Andererseits wird vermutet, dass die Umdeutung des Fußballrhythmus aufgrund seiner Bekanntheit „nicht wirklich funktioniert“ (*Lk.18_TA.aa_Vb*). Zudem ergeben sich durch die Verfremdung des Rhythmus' Überbindungen einzelner Noten, die für das Verstehen von Schüler*innen als problematisch erachtet werden (z. B. *Hsl.6_TA.aa_Bg*; *Stud.49_TA.a_Rf*). Beim Herstellen von Lebensweltbezug durch bekannte Musikbeispiele muss demnach bedacht werden, inwiefern diese für methodisch-didaktische Überlegungen adaptiert werden können.

Auch wird kritisiert, dass ein bloßes Klatschen und Stampfen nicht dem tagtäglichen Musikerleben von Schüler*innen entspricht und es der Erklärung dadurch an Lebensweltbezug fehle (*Hsl.4_TA.aa_Rf*). Der Vergleich zwischen den Verhältnissen von Notenwerten gegenüber der Aufteilung einer Zeitstunde, wird mit Blick auf den Lebensweltbezug der Erklärung hauptsächlich positiv bewertet (z. B. *Lk.13_NW.o_Bg*; *Stud.62_NW.o_Bg*). Allerdings gibt es auch in diesem Kontext Einschränkungen bezüglich der Frage, wie gut die Analogie zwischen der Aufteilung von Notenwerten und einer Zeitstunde passt.¹⁴⁸

Die Orientierung am Vorwissen der Schüler*innen wird auf unterschiedlichen Ebenen thematisiert. Zum einen gibt es enge Überschneidungen zum Lebensweltbezug, indem der Fußballrhythmus auch an das „Vorwissen der Lernenden“ (*Hsl.8_TA.aa_Bg*) andockt. Dass hier Vorwissen im Sinne von deklarativem Wissen als passender Begriff dienlich ist, kann jedoch bezweifelt werden. Vergleichbar positiv wird der Einsatz des Liedes *Dracula Rock* bewertet, dem eine ähnliche Form der Bekanntheit attestiert wird (*Hsl.4_Tl.a_Bg*), da in den Erklärungen zum Thema Tonleiter erwähnt wird, dass es in einer vorherigen Unterrichtsstunde gesungen wurde. Zum anderen wird fehlendes Vorwissen im Kontext deklarativer Wissensinhalte vermutet. Dabei sind Unterschiede zwischen den beiden Statusgruppen auffällig: Während Studierende fehlendes Vorwissen der Schüler*innen zu konkreten Inhalten wie Intervallen, dem Unterschied von Ganz- und Halbtonschritten oder dem Aufbau einer Dur-Tonleiter vermuten (z. B. *Stud.40_TA.aa_Rf*; *Stud.9_TL.aa_Bg*; *Stud.32_TL.aa_Bg*), bleiben die Lehrenden in ihren Aussagen eher vage bezüglich konkreter Inhalte (z. B. *Lk.19_TL.o_Vb*; *Hsl.9_NW.aa_Bg*); befürchtet wird nur, dass nicht alle Schüler*innen über ausreichendes Vorwissen verfügen.

Ein weiterer, dem Vorwissen nahestehender Aspekt, richtet sich auf die Berücksichtigung unterschiedlicher musikalischer Vorerfahrung von Schüler*innen. Unterschieden wird zwischen Lernenden, die beispielsweise aufgrund außerschulischen Instrumental- oder Gesangsunterrichts über ausgeprägte Fähigkeiten verfügen, und solchen, denen diese fehlen (z. B. *Lk.3_NW.aa_Bg*). Tendenziell werden dabei akustische Repräsentationen insbesondere für die Schüler*innen als hilfreich erachtet, die im Allgemeinen bereits über ausgeprägte musikalische Vorerfahrung (z. B. *Lk.16_TL.a_Bg*; *Stud.9_TA.a_Bg*)

148 Der Einsatz von Analogien wird im weiteren Verlauf des Kapitels gesondert besprochen. Im Gegensatz zu Aspekten des Lebensweltbezugs wird die Passung der Analogie zwischen Notenwerten und Stunde mehr kritisiert.

beziehungsweise im Speziellen über relevante Hörkompetenzen verfügen (z. B. *Hsl.8_TL.a_Bg*; *Stud.39_TL.a_Bg*; *Slk.15_Tl.a_Bg*).

Hingegen wird eine Erklärung, die nur mit akustischer und ohne optische Repräsentation auskommt, für leistungsschwache Schüler*innen als zu schwierig empfunden (*Hsl.11_TA.a_Rf*); für diese Gruppe lautet die nachdrückliche Empfehlung an vielen Stellen, optische Repräsentationen einzusetzen (z. B. *Stud.27_TL.oa_Bg*; *Lk.23_TL_a_Bg*; *Stud.5_TL.a_Bg*). Vereinzelt werden jedoch auch gegensätzliche Meinungen vertreten: Explizit gewürdigt wird an einer Stelle das Weglassen von Noten, da diese „die meisten SuS [*Schülerinnen und Schüler*; vom Autor hinzugefügt] vermutlich doch eher überfordern“ (*Prof.8_TL.a_Bg*). Aus einer anderen Perspektive wird bemängelt, dass die Verwendung der Klaviatur eine Bevorzugung derjenigen darstellt, die selbst Klavier spielen und dadurch über eine konkrete musikalische Vorerfahrung verfügen (*Slk.9_TL.o_Bg*).

Zudem sprechen die Studienteilnehmer*innen mögliche Missverständnisse oder daraus resultierende Fehlvorstellungen an, die aufgrund der Erklärungen bei Schüler*innen ausgelöst werden können. Beispielsweise suggeriere die Äußerung, wie lange ein Ton klingt, „dass es um absolute Tondauern gehen könnte“ (*Hsl.4_NW.oa_Rf*; siehe auch *Prof.1_NW.oa_Bg*). Vorgeschlagen wird in diesem Fall die zusätzliche Erläuterung, dass zwar die Relationen der Notenwerte immer gleich sind, deren absolute Länge jedoch abhängig vom gewählten Tempo ist (*Stud.27_NW.oa_Bg*).

Auch besteht bei der Darstellung von Prozessschritten die Gefahr, dass diese nicht verallgemeinert werden können, wenn die Umsetzung mit Hilfe eines konkreten Beispiels erfolgt und darauf nicht explizit hingewiesen wird. Das Herauslösen der Töne aus dem Lied *Dracula Rock* führt dazu, dass die Töne zwei und drei sowie vier und fünf dieser Reihe getauscht werden müssen, um eine stufenweise Abfolge der Töne in einer Tonleiter zu erhalten; fehlt ein Hinweis auf die Beispielhaftigkeit, könnten Schüler*innen denken, dass immer diese Töne getauscht werden müssen (*Lk.22_TL.o_Bg*). Auch wird in diesem Kontext ein anderes mögliches Missverständnis vermutet: „Die Erklärung könnte bei den Schülern die Vorstellung erzeugen, eine Tonleiter wäre ästhetisch hochwertiger als die ungeordnete Melodie“ (*Slk.14_TL.oa_Vb*).

Zusammenfassend stützen die Aussagen einerseits die Relevanz des Qualitätsmerkmals *Adressatenorientierung*, wie sie bereits für die Gruppen der Studierenden und Lehrenden in den Regressionsmodellen inferenzstatistisch nachgewiesen werden konnte. Andererseits differenzieren sie auch mögliche Facetten dieses Merkmals aus. Für unterrichtliches Erklären im Fach Musik kann demnach weiterhin angenommen werden, was bereits vielfach diskutiert wird (z. B. Linn, 2017): Die Heterogenität der Schüler*innen hinsichtlich deren musikalischer Erfahrung ist beim Erklären zu berücksichtigen. Gerade für den Einsatz akustischer Repräsentationen in Form von Hör- und Klangbeispielen erscheint es notwendig, genau abzuwägen, inwieweit sich diese für einen Teil der Schüler*innen zwar positiv, für andere aber auch negativ – beispielsweise im Sinne einer Überforderung – auswirken könnte. Für Lernende mit wenig musikalischer Vorerfahrung werden insbesondere optische Repräsentationen beim Erklären als notwendig erachtet.

Diese beiden Beobachtungen decken sich mit den Unterschieden in der Bewertung der Erklärqualität (siehe Kap. 5.3.3): Neben Schüler*innen bewerten auch Studierende

und Lehrende in der Tendenz die Erklärungen am schlechtesten, in denen eine optische Repräsentation fehlt, während das Fehlen einer akustischen zwar deskriptiv etwas schlechter bewertet wird, jedoch zu keinen signifikanten Unterschieden führt. Die Berücksichtigung des Lebensweltbezugs der Schüler*innen wird für das unterrichtliche Erklären grundsätzlich positiv bewertet. Allerdings ist dabei immer zu prüfen, ob aus dem Alltag bekannte Melodien, Rhythmen oder Konzepte wie die Aufteilung einer Zeitstunde auch unter methodisch-didaktischen Veränderungen wie der Verfremdung eines Rhythmus' sinnvoll eingesetzt werden können. Lebensweltbezug darf somit kein Selbstzweck sein.

Strukturiertheit

Zu *Strukturiertheit* liegen 198 codierte Stellen vor, von denen 130 als Begründung zum Globalurteil abgegeben wurden. Zwar entfallen damit im Vergleich zu *Adressatenorientierung* insgesamt etwas weniger codierte Segmente auf dieses Merkmal, dennoch wird angenommen, dass die Studienteilnehmer*innen auch die *Strukturiertheit* einer Erklärung als Qualitätsmerkmal ansehen. Auch hier liegen inhaltliche Überschneidungen zwischen offenen Antworten und den Skalenitems der Operationalisierung von *Strukturiertheit* vor, jedoch sind in den Aussagen deutliche Schwerpunkte zu erkennen.

Die häufigsten Aussagen thematisieren den Aufbau der Erklärung. Allerdings sind dabei Unterschiede zwischen den drei Themen festzustellen. Für die Erklärungen zu Notenwerten wurde nicht nur kritisiert, dass mit der halben Note begonnen wurde (z. B. *Prof.4_NW.oa_Bg*; *Prof.6_NW.oa_Vb*), sondern in den meisten Fällen auch empfohlen, entweder mit der Viertelnote oder der ganzen Note in das Thema Notenwerte einzusteigen (z. B. *Hsl.8_NW.oa_Bg*; *Lk.18_NW.oa_Bg*; *Lk.3_NW.oa_Vb*; *Prof.8_NW.o_Vb*). Diese Vorschläge zielen auf einen inhaltlich-logischen Aufbau, der die Reihenfolge von Groß nach Klein beziehungsweise Klein nach Groß gegenüber einer Entwicklung aus der Mitte in verschiedene Richtungen bevorzugt.

Ein weiterer Aspekt bezieht sich auf die Frage, wann in einer Erklärung Musik erklaren sollte. Dieser kann anhand der Aussagen nicht einheitlich beantwortet werden. So kristallisiert sich aus den Aussagen der Studierenden und Lehrenden zum Thema Notenwerte heraus, dass Begriffe wie *Schlag*, denen ein bestimmtes Konzept zugrunde liegt, zu Beginn erklärt werden sollten (z. B. *Hsl.11_NW.oa_Bg*; *Stud.39_NW.oa_Bg*). Klangliche Demonstrationen sollten demnach erst erfolgen, nachdem eine Übersicht gegeben oder ein Konzept erklärt wurde (z. B. *Stud.7_NW.oa_Bg*; *Stud.10_NW.oa_Bg*). So lobt eine Lehrkraft den „Praxisbezug mittels Klavier“, sieht diesen jedoch „an der falschen Stelle“ und erachtet eine akustische Repräsentation erst dann als sinnvoll, „wenn alle Notenwerte in der Übersicht aufgeführt sind“ (*Lk.23_NW.oa_Bg*).

In eine ähnliche Richtung zielen Aussagen zu den Erklärungen des Themas Tonleiter. So folgt der Vorschlag, zunächst das Prinzip der Halbtonabstände zu klären, um Schüler*innen zur eigenständigen Erschließung des Aufbaus einer Tonleiter zu befähigen (*Slk.3_TL.o_Vb*), der Idee, durch kurze Erklärungen konstruktivistisches Lernen zu ermöglichen und die Schüler*innen dadurch auch zu motivieren (*Stud.60_TL.o_Vb*). Auch

wird insbesondere mit Blick auf Nicht-Musiker*innen empfohlen, „zunächst das System [der Ganz- und Halbtönschritte in einer Tonleiter, vom Autor hinzugefügt] zu klären und es anschließend anzuwenden“ (Slk.9_TL.o_Bg).

Im Gegensatz dazu wird vielfach auch der Wunsch geäußert, das Lied *Dracula Rock* als Einstieg in die Erklärung mit der Klasse zu singen oder den Schüler*innen vorzuspielen (z. B. Stud.44_TL.o_Vb; Slk.3_TL.o_Vb; Lk.17_TL.a_Bg; Hsl.3_TL.oa_Vb). Auffällig ist dabei, dass für diesen Wunsch keine didaktischen Gründe angeführt werden. So wäre in diesem Zusammenhang auch denkbar, dass das einleitende Musizieren hauptsächlich zum Zweck der Motivation erfolgt. Auch wird positiv bewertet, wenn zuerst eine akustische Demonstration des musikalischen Phänomens erfolgt, welche dann jedoch direkt im Anschluss durch eine optische Darstellung ergänzt wird (z. B. Hsl.5_TA.oa_Bg; Hsl.8_TA.a_Bg; Stud.60_TA.a_Bg). Eine solche Visualisierung kann dabei auch zum Zweck einer Zusammenfassung am Ende dienlich sein – beispielsweise um die drei besprochenen Taktarten wiederum gegenüberzustellen und die relevanten Unterschiede zu verdeutlichen (z. B. Hsl.8_TA.oa_Bg; Slk.12_TA.oa_Bg).

Neben dem Aufbau wird in den offenen Antworten auch als wichtig erachtet, inwieweit sich eine Erklärung auf wesentliche Inhalte fokussiert. So werden bei der Erklärung NW_oa überflüssige Informationen, wie die Zeitstunden-Analogie oder der Hinweis, dass die halbe Note ihren Namen nicht habe, weil sie etwa kaputt sei, kritisch angemerkt (Hsl.2_NW.oa_Bg; Lk.23_NW.oa_Bg; Stud.50_NW.oa_Bg). Für andere Studienteilnehmer*innen hingegen liegen die wesentlichen Informationen bei dieser Erklärung vor (z. B. Hsl.8_NW.oa_Bg; Stud.43_NW.oa_Bg). Bei den Erklärungen zum Thema Tonleiter wurde beispielsweise kritisiert, dass auf den Aspekt Chromatik eingegangen wurde, der jedoch für eine Dur-Tonleiter nicht vorrangig relevant ist (Hsl.4_TL.a_Vb; Stud.8_TL.a_Vb; Stud.23_TL.o_Bg). Auch wurde bemängelt, dass der Unterschied von Ganz- und Halbtönschritten bei der Erklärung einer Dur-Tonleiter eine „Vermengung von verschiedenen Inhalten“ (Slk.9_TL.a_Bg) darstellt; zudem sollte der Unterschied in der 6. Jahrgangsstufe bekannt sein (Slk.9_TL.o_Bg).

Weitere Facetten der *Strukturiertheit* einer Erklärung werden in den offenen Antworten nur vereinzelt thematisiert und sollen die Darstellung des Merkmals abschließen. Im Sinne der Kohärenz einer Erklärung wird positiv bewertet, wenn auf den Beginn einer Erklärung am Ende zurückgegriffen und ein Bezug hergestellt wurde (z. B. beim Thema Tonleiter die Verwendung von Legosteinen als Analogie für Tonmaterial; siehe Stud.23_TL.a_Bg; Stud.8_TL.oa_Bg). Es liegen kaum Aussagen dazu vor, inwieweit das Ziel der Erklärung klar war (z. B. Lk.3_TA.oa_Rf) – dieser Aspekt wurde in die Operationalisierung der Skala als Item aufgenommen.

Zusammenfassend können die offenen Antworten die *Strukturiertheit* einer Erklärung als Qualitätsmerkmal stützen. Für die Gruppe der Studierenden zeigte sich dieser Effekt bereits in den inferenzstatistischen Analysen (siehe Tab. 5.36). Da im Vergleich dazu die *Strukturiertheit* für die Lehrenden jedoch keinen signifikanten Einfluss auf die Bewertung der Erklärqualität hatte, wäre es naheliegend, auch in den offenen Antworten Unterschiede zwischen Studierenden und Lehrenden zu sehen. Diese können jedoch weder hinsichtlich der Anzahl an Äußerungen (96 Codierungen bei Studierenden, 102 Codierungen

bei Lehrenden) noch bezüglich einzelner Facetten in den Aussagen der Studienteilnehmer*innen festgestellt werden.

Inhaltlich kristallisieren sich insbesondere zwei Facetten von *Strukturiertheit* heraus. Zum einen wird die akustische Darstellung von Musik an verschiedenen Positionen einer Erklärung vorgeschlagen. Dabei wird das Hören oder Singen von Musik grundsätzlich als Einstieg in eine Erklärung befürwortet – die expliziten Zwecke werden in den offenen Antworten aber nicht thematisiert, motivationale Ziele sind jedoch denkbar. Im Kontext kognitiver Inhalte von Musikunterricht kann hingegen eher die Tendenz festgestellt werden, zunächst grundlegende Prinzipien und Konzepte zu erklären und eine akustische Darstellung im Anschluss zu präsentieren. Die Erklärung solcher Prinzipien sollte zum anderen einem inhaltlich-logischen Aufbau folgen. Zudem sollten nach Aussage der Studienteilnehmer*innen bei einer Erklärung im Fach Musik wesentliche Inhalte fokussiert werden. Andere Aspekte der *Strukturiertheit* einer Erklärung wie deren Kohärenz sowie die Verdeutlichung der Erklärziele werden hingegen nur vereinzelt genannt.

Sprachliche Verständlichkeit

Auf die Kategorie Sprache entfallen 159 Codierungen, von denen 99 aus Aussagen zur Begründung der Globalbewertung stammen. Obwohl die Häufigkeit der Nennung im Vergleich zu *Adressatenorientierung* und *Strukturiertheit* nochmals abgeschwächt ist, wird die Bedeutsamkeit sprachlicher Aspekte als Qualitätsmerkmal angenommen. Inhaltlich liegen neben allgemeinen Bewertungen zu *Sprachliche Verständlichkeit* vor allem Aussagen zur Verwendung von (Fach-)Begriffen vor. Während sich diese beiden Aspekte inhaltlich mit der Operationalisierung der Itemskala überschneiden, werden in den offenen Antworten weitere Aspekte thematisiert, die jedoch deutlich seltener angesprochen werden: Eine Rolle spielt zum einen die Menge an gesprochenem Text, zum anderen werden fachspezifische Formulierungen hinsichtlich ihrer Verständlichkeit für Schüler*innen beurteilt.

Mit Blick auf die in den Erklärungen verwendeten (Fach-)Begriffe liegen unterschiedliche Meinungen vor. Während für die Erklärungen zu Taktarten tendenziell positiv bewertet wird, wie die Begriffe Puls, Takt, Betonung oder Rhythmus voneinander abgegrenzt und in Beziehung gesetzt wurden (z. B. *Lk.11_TA.0a_Bg*; *Hsl.4_TA.a_Bg*; *Stud.60_TA.0a_Bg*), wird auch moniert, dass dies zu viele Begriffe seien (*Slk.10_TA.0a_Bg*) oder andere Worte wie „Rhythmus“ (*Slk.8_TA.0a_Vb*) oder „Oktave“ (*Stud.58_TL.0a_Vb*) nicht ausreichend erklärt wurden.

In diesem Zusammenhang wird auch als wichtig erachtet, dass Begriffe nicht nur verbal erklärt, sondern auch visualisiert werden (*Prof.3_TL.a_Bg*; *Hsl.8_TL.a_Vb*) – unklar bleibt in den Aussagen jedoch, ob die Visualisierung in Form von Notentext, der Niederschrift des Begriffs, einer Kombination aus beidem oder weiteren Aspekten präsentiert werden sollte. Häufig kritisiert wird insbesondere die Verwendung des Begriffs *Schlag*. Dieser ist als gängiges Synonym für *Puls* beziehungsweise als Verkürzung des Wortes *Grunds Schlag* zu verstehen, wird jedoch mit Blick auf das Verstehen von Schüler*innen

ohne eine ausführliche Erklärung als missverständlich und abstrakt eingeschätzt (z. B. *Lk.16_NW.0a_Bg*; *Slk.1_NW.0a_Bg*).

Weniger häufig nehmen die Aussagen Bezug auf fachtypische und -spezifische Formulierungen. Dabei wird es beispielsweise für Ungeübte als zu abstrakt bewertet, wenn von ganzen und halben Schritten sowie Tönen, die „dazwischen passen“ (*Slk.15_TL.0_Bg*; siehe auch *LK.18_TL.0_Bg*) gesprochen wird. Auch seien Formulierungen wie „Das Cis ist stumm“ (*Slk.23_TL.0a_Rf*) oder „versteckter Ton“ (*Lk.3_TL.a_Bg*) problematisch und könnten für Schüler*innen mitunter schwer verständlich sein. Auch wird die Formulierung „Puls oder Takt darunter denken“ (*Lk.15_TA.a_Bg*) als unklar bewertet.

Ein weiterer, ebenfalls nur vereinzelt auftretender Aspekt, stellt die Menge an gesprochenem Text dar. Da zu jeder Erklärung mindestens eine Äußerung dazu vorliegt, dass beispielsweise „nicht zu lange am Stück“ (*Lk.24_TL.0a_Vb*) gesprochen werden soll oder die „Tendenz zu Redefluss“ (*Slk.10_TL.0_Bg*) vorliegt, erscheint dies durchaus relevant für *Sprachliche Verständlichkeit* zu sein. Eine Interpretation hinsichtlich möglicher Zusammenhänge mit dem Erklärmodus oder einzelnen Themen erscheint bei insgesamt nur zwölf Codierungen jedoch zu wenig aussagekräftig.

Zusammenfassend ist auf Basis der Daten festzuhalten, dass insbesondere die Verwendung von Fachbegriffen eine bedeutende Rolle für die sprachliche Verständlichkeit einer Erklärung spielt. Eine weitere bedeutsame Facette könnte die achtsame Verwendung fachtypischer, jedoch sprachlich missverständlicher Formulierungen darstellen. Zudem zeigt sich die Menge an gesprochenem Text als potenzielle Facette von *Sprachliche Verständlichkeit*. Nicht thematisiert wurde in den offenen Antworten die Satzlänge, die als weitere Aspekte in die Operationalisierung eingeflossen ist. Dieses Ungleichgewicht könne ein Grund für die in den Regressionsmodellen teils schwach, teils negativ ausgeprägten Effekte bei *Sprachliche Verständlichkeit* sein. Zudem kann eine sprachwissenschaftliche Analyse, die für alle beteiligten Fächer in FALKE-q erfolgt ist (siehe Thim-Mabrey & Lindl, i. Vorb.), weiteren Aufschluss über den Aspekt *Sprachliche Verständlichkeit* unterrichtlichen Erklärens im Fach Musik geben.

Sprech- und Körperausdruck

Im Vergleich zu den bisherigen drei Kategorien liegen für das Merkmal *Sprech- und Körperausdruck* insgesamt nur 49 Codierungen vor. Da von diesen lediglich 22 Aussagen auf die Begründung des Globalurteils entfallen, kann auf Basis dieser Aussagen in den offenen Antworten nicht von einer Bedeutung für die Qualität von Erklärungen ausgegangen werden. Das deckt sich auch mit den inferenzstatistischen Befunden der Regressionsmodelle: Weder bei den Studierenden noch bei den Lehrenden konnte ein signifikanter Zusammenhang mit der Globalbewertung der Erklärung festgestellt werden.

Die Urteile für die sprecherischen Aspekte fallen größtenteils positiv aus. Sie beziehen sich insbesondere auf das Sprechtempo (*Hsl.1_TL.0_Bg*; *Stud.62_TL.0a_Rf*; *Lk.23_NW.0_Rf*) und nur vereinzelt auf eine deutliche Aussprache (*Lk.15_NW.0_Bg*), eine angenehme Sprechweise (*Stud.43_TA.a_Rf*) und auf Sprechpausen (*Hsl.8_TA.a_Rf*).

Vereinzelt sind jedoch auch gegensätzliche Wahrnehmungen festzustellen: Während der Vortragsstil in Erklärung NW_o einerseits als „rhetorisch/sprachlich einwandfrei“ (Lk.23_NW.o_Rf) bezeichnet wird, schlägt eine andere Person vor, bei dieser Erklärung flüssiger zu sprechen und bezeichnet den Vortrag als „etwas stockend“ (Slk.4_NW.o_Vb).

Kaum eine Rolle spielt in den offenen Antworten der Blickkontakt der Lehrkraft. Während die sprechsprachlichen Elemente der Operationalisierung von *Sprech- und Körperausdruck* zumindest vereinzelt genannt werden, spielt dieser auf die Körpersprache abzielende Aspekt keine Rolle. Eine andere Facette von Körpersprache wiederum wird thematisiert: Mehrfach wurde moniert, dass die Lehrkraft beim Beschreiben von Notenkopf und Notenhals in der Erklärung NW_oa von einem ausgefüllten Notenkopf spricht und dabei auf eine Note zeigt, deren Notenkopf nicht ausgefüllt ist (z. B. Hsl.12_NW.oa_Bg; Slk.7_NW.oa_Bg; Lk.24_NW.oa_Bg; Stud.3_NW.oa_Bg).

Hinsichtlich der Kontrollvariable *Persönlichkeitswirkung*, die eine inhaltliche Nähe zu sprecherischen und körpersprachlichen Aspekten aufweist, können in den offenen Antworten keine nennenswerten Aussagen festgestellt werden. Dieser Befund unterstützt die Ergebnisse der Regressionsanalysen, in denen die *Persönlichkeitswirkung* der Lehrkraft keinen signifikanten Einfluss auf die Bewertung der Erklärqualität einnimmt.

Zusammenfassend zeigt die Analyse der offenen Antworten für Studierende und Lehrende, dass für die Studienteilnehmer*innen der *Sprech- und Körperausdruck* der Lehrkraft keine bedeutende Rolle mit Blick auf die Erklärqualität spielt. Gegenüber den anderen Merkmalen liegt nicht nur eine deutlich geringere Zahl an Aussagen in den offenen Antworten vor. Zudem können nur für die Aspekte Gestik (beim Zeigen auf den Notenkopf in Erklärung NW_oa) und für das Sprechtempo der Lehrkraft Äußerungen mehrerer Personen festgestellt werden. In Kombination mit den Ergebnissen der Regressionsanalysen muss daher für diese beiden Statusgruppen infrage gestellt werden, ob *Sprech- und Körperausdruck* ein Qualitätsmerkmal unterrichtlichen Erklärens im Fach Musik darstellt. Aufgrund der fehlenden Daten der Schüler*innen bleibt diesbezüglich offen, ob dies auch für diese Statusgruppe gelten kann, da *Sprech- und Körperausdruck* im linearen Modell den größten Einfluss gezeigt hat.

Einsatz musikbezogener Repräsentationen

Zum *Einsatz musikbezogener Repräsentationen* wurden insgesamt 760 Textstellen codiert. Davon entfallen 453 Codierungen auf offene Antworten, anhand derer die Bewertung der Erklärqualität durch Studierende (200 Textstellen) und Lehrende (253 Textstellen) annähernd gleich häufig begründet wurde. Aufgrund dieser hohen Zahl wird angenommen, dass der *Einsatz musikbezogener Repräsentationen* ein bedeutsames Qualitätsmerkmal unterrichtlichen Erklärens darstellt. Da für diesen Prädiktor jedoch weder im Regressionsmodell der Studierenden (siehe Tab. 5.37) noch in dem der Lehrenden (siehe Tab. 5.38) ein signifikanter Einfluss nachgewiesen werden kann, übernimmt die Analyse der offenen Antworten zweierlei Aufgabe: Zum einen wird erwartet, dass anhand der Auswertungen Erkenntnisse über Facetten des Merkmals *Einsatz musikbezogener*

Repräsentationen gewonnen werden können. Zum anderen stellen diese Befunde wichtige Anhaltspunkte für eine zukünftige Operationalisierung des Merkmals dar.

Aufgrund der Fülle an Codierungen erfolgt der Bericht getrennt nach vier verschiedenen Formen der Repräsentation: Die beiden größten und umfangreichsten stellen optische und akustische Repräsentationen dar. Daneben wird auch der Einsatz von Analogien und Vergleichen in den Aussagen thematisiert. Die vierte Form der Verwendung von Repräsentationen bezieht sich auf konkrete (Musik-)Beispiele, die jedoch nahezu ausschließlich im Kontext der Erklärungen zu Taktarten abgegeben werden. Im Vergleich zu den Textstellen, in denen Überzeugungen zum Einsatz der Repräsentationsformen zum Tragen gekommen sind und die zur Beantwortung von F4c (siehe Kap. 3.2.2) herangezogen wurden, liegen der folgenden Auswertung Äußerungen zugrunde, die sich auf den Einsatz einer Repräsentation und deren konkrete Umsetzung in der Erklärung beziehen.

Optische Repräsentationen

Analog zu den bisherigen Qualitätsmerkmalen liegen auch für optische Repräsentationen allgemeine Bewertungen wie „gelungene Visualisierung“ (*Slk.9_TA.0a_Bg*), „schlüssige optische Darstellung“ (*Prof.6_TL.0_Bg*) oder „anschaulich visualisiert“ (*Hsl.4_TL.0_Bg*) vor. Die codierten Textstellen verteilen sich dabei über alle fünf Erklärungen, in denen optische Repräsentationen eingesetzt wurden. Daneben können insbesondere zwei Aspekte herausgearbeitet werden, die vielfach thematisiert wurden und mitunter kontroverse Positionen erkennen lassen: Einerseits spielt die Darstellung von Notentext eine besondere Rolle. Andererseits wird der Einsatz einer Klaviatur als Hilfsmittel thematisiert, dabei jedoch unterschiedlich bewertet.

Die Studienteilnehmer*innen schreiben der Abbildung von Notentext unabhängig einzelner Themen eine wichtige Rolle zu. Das Zeigen von Notentext kann nicht nur dazu dienen, das mündlich Erklärte am Notenmaterial zu verifizieren (*Slk.12_TL.a_Bg*), sondern auch dazu, Schüler*innen beim Verstehen und Nachvollziehen der Inhalte zu helfen (z. B. *Prof.1_TL.a_Bg*; *Slk.4_TA.a_Bg*; *Lk.23_TA.a_Bg*; *Stud.62_TA.a_Vb*). So seien die Verschiebungen des Rhythmus in Erklärung TA_a, die sich aus den verschiedenen Taktarten ergeben, „ohne Notenbeispiel nur schwer verständlich“ (*Slk.4_TA.a_Bg*). Hervorgehoben werden auch Visualisierungen im Notentext, die beispielsweise Taktschwerpunkte farblich markieren (z. B. *Hsl.2_TA.0a_Bg*; *Stud.58_TA.0a_Rf*) oder die Tondauer in Form von Zeitbalken abbilden (z. B. *Hsl.11_NW.0a_Rf*; *Stud.23_NW.0a_Bg*; *Stud.18_NW.0a_Rf*). Etwas überraschend wird die Darstellung von Notentext für Schüler*innen auch als verwirrend und überfordernd bewertet: So wird vor allem bei der Erklärung TA_a, bei der derselbe Rhythmus in verschiedenen Taktarten akustisch repräsentiert wird, in manchen Aussagen positiv hervorgehoben, dass keine Visualisierung von Notentext stattfindet (z. B. *Lk.3_TA.a_Bg*; *Lk.20_TA.a_Bg*).

Treten Visualisierungen als Hilfsmittel für ein besseres Verstehen in Erscheinung, werden vor allem inhaltliche und methodische Fehler kritisiert. So wird die Aufteilung der Notenwerte in Form einer Pyramide als ungünstig bewertet, da sowohl der horizontale Verlauf von links nach rechts sowie der vertikale Verlauf, durch den der gemeinsame Startpunkt der Töne verdeutlicht werden kann, verloren gehen (z. B. *Prof.3_NW.0_Bg*;

Slk.8_NW.o_Bg; Slk.1_NW.oa_Vb). Weiterhin wird angemerkt, dass die Schüler*innen durch die Effekte der Visualisierung abgelenkt werden (z. B. *Lk.3_TA.aa_Rf; Lk.17_NW.aa_Bg; Hsl.11_NW.o_Rf*).

Hilfreich können Visualisierungen für eine Zusammenfassung und eine knappe Übersicht sein; beispielsweise wird für die Erklärung *TA_oo* als Verbesserung vorgeschlagen, die drei verschiedenen Taktarten am Ende als Überblick gemeinsam zu zeigen und anhand dessen eine kurze Zusammenfassung vorzunehmen (z. B. *Hsl.8_TA.aa_Bg; Slk.12_TA.aa_Bg*). Hier zeigt sich eine Verbindung zum Merkmal Strukturiertheit, das als wichtige Facette eine Zusammenfassung am Ende der Erklärung vorsieht. Auch wird es als notwendig erachtet, Visualisierungen zu erklären: So sollte das Aussehen der Notenwerte kurz erklärt werden (z. B. *Slk.9_NW.o_Bg; Hsl.12_NW.aa_Bg*). Zudem erfordert der Einsatz einer Klaviatur eine Erklärung, wenn sie als Strukturierungshilfe für Ganz- und Halbtonschritte verwendet wird (*Slk.14_TL.o_Vb*).

Die Verwendung der Klaviatur im Kontext der Erklärungen zu Tonleitern wird in den Aussagen kontrovers diskutiert: Auf der einen Seite wird der Einsatz mit Blick auf das Verstehen der Schüler*innen als hilfreich bewertet, um die Abfolge von Halb- und Ganztonschritten in einer Tonleiter zu erklären (z. B. *Slk.4_TL.o_Bg; Hsl.2_TL.o_Bg; Prof.4_TL.o_Bg; Stud.37_TL.o_Bg; Slk.15_TL.aa_Bg*). Allerdings werden gleichermaßen auch Limitationen und Einschränkungen dieser Visualisierungshilfe genannt: Aufgrund der ungleichen Anordnung der Halbtöne mittels weißer und schwarzer Tasten wird das Gitarrengriffbrett anstelle der Klaviatur vorgeschlagen, da alle Bünde denselben Abstand haben und demnach keine bautechnischen Besonderheiten zu erklären sind (z. B. *Slk.4_TL.aa_Rf; Slk.12_TL.aa_Bg; Lk.24_TL.aa_Bg*).

Insbesondere bei Erklärung *TL_a*, also der Erklärung ohne eine optische Repräsentation, wird die Forderung, eine Tonleiter auch in Form von Notentext abzubilden (z. B. *Prof.1_TL.a_Bg; Lk.15_TL.a_Bg*), häufig mit dem Wunsch nach einer Klaviatur als Visualisierungshilfe verknüpft (z. B. *Hsl.1_TL.a_Vb; Lk.3_TL.a_Bg; Lk.18_TL.a_Vb*). Die ambivalente Haltung gegenüber der Klaviatur als Hilfsmittel zur Verdeutlichung von Skalenstrukturen fasst folgendes Zitat zusammen: Deren Einsatz sei „so wirksam wie fragwürdig spezifisch“, weshalb das Zirkelmodell [*gemeint ist vermutlich das von Beiderwieden (2008)*, vom Autor hinzugefügt] empfohlen und mit Blick auf die Lernenden als „neutraler im Hinblick auf einen unterstellten oder nicht unterstellten instrumentalen Hintergrund“ (*Slk.14_TL.aa_Vb*) bewertet wird.

Zusammenfassend können die Erkenntnisse aus der *mixed ANOVA* (siehe Kap. 5.3) anhand der inhaltsanalytischen Auswertung unterstützt werden. Das Fehlen einer optischen Repräsentation fiel dort signifikant negativ ins Gewicht. Die offenen Antworten der Studierenden und Lehrenden können dahingehend insbesondere die Darstellung von Notenmaterial als Notwendigkeit spezifizieren. Unterstützende Markierungen im Notentext können ebenso zum Verständnis der Schüler*innen beitragen wie eine zusammenfassende Darstellung der relevanten Aspekte am Ende einer Erklärung. Der Einsatz einer Klaviatur als Visualisierungshilfe wird hinsichtlich des Verstehens der Schüler*innen einerseits positiv gewürdigt, andererseits als Pianist*innen bevorzugend kritisiert.

Akustische Repräsentationen

Gegenüber den optischen liegen zu den akustischen Repräsentationen deutlich weniger allgemeine Bewertungen wie „gut klanglich dargestellt“ (*Hsl.9_TA.0a_Bg*) oder „auch gut am Klavier demonstriert“ (*Lk.2_TL.0a_Bg*) vor. Daneben beziehen sich die Aussagen hauptsächlich darauf, inwieweit akustische Darstellungen zum Verstehen der Schüler*innen beitragen. Deutlich seltener werden jedoch auch die Verschiedenartigkeit der Hörbeispiele, die Auswahl an konkreten Musikbeispielen sowie die Häufigkeit des Vorspielens beziehungsweise Vormachens von Hör- und Klangbeispielen thematisiert.

Mit Blick auf die Frage, wie der Einsatz von akustischen Repräsentationen das Verstehen der Schüler*innen unterstützen kann, lohnt bei der Auswertung der Aussagen eine Darstellung, die getrennt nach den jeweils verwendeten Erklärmodi erfolgt. So ist bei beiden Erklärungen ohne akustische Repräsentation (*NW_o*; *TL_o*) auffällig, dass der Wunsch nach einer solchen eher spärlich ausfällt. Aussagen wie „Ganz- und Halbtöne durch Hörbeispiele verdeutlichen“ (*Hsl.8_TL.o_Vb*) oder „ein Notenbeispiel mit den verschiedenen Notenwerten am Klavier spielen“ (*Stud.13_NW.o_Vb*) sind nur vereinzelt zu finden. Dabei treten jedoch vielfältige Vorteile einer klanglichen Darstellung hervor: Das Vorspielen der Halbtöne am Klavier trage demnach sowohl zur Auflockerung, zur Veranschaulichung des Inhalts als auch zu einer besseren klanglichen Vorstellung bei (z. B. *Lk.2_TL.o_Bg*; *Lk.15_TL.o_Bg*; *Stud.43_TL.o_Bg*).

Im Gegensatz dazu unterscheiden sich die Aussagen zum Einsatz akustischer Repräsentationen bei den beiden Erklärungen, die ohne optische Repräsentation erfolgen, deutlich: So vermuten die Studienteilnehmer*innen bei der entsprechenden Erklärung zur Tonleiter (*TL_a*), dass ein Großteil der Schüler*innen Ganz- und Halbtöne kaum unterscheiden und der Erklärung dadurch nicht folgen kann (z. B. *Lk.3_TL.a_Bg*; *Lk.8_TL.a_Bg*; *Lk.9_TL.a_Bg*; *Lk.11_TL.a_Vb*; *Slk.15_TL.a_Bg*; *Hsl.2_TL.a_Bg*; *Hsl.12_TL.a_Bg*; *Stud.4_TL.a_Bg*; *Stud.41_TL.a_Bg*) – dies könnte ein entscheidender Grund dafür sein, dass diese Erklärung von allen Statusgruppen im Schnitt am schlechtesten bewertet wurde (siehe Tab. 5.14).

Angemerkt wird außerdem, dass „der Verzicht auf graphische Erläuterung (...) als Ausgleich eine intensivere, vielschichtigere akustische Verdeutlichung erfordert“ (*Slk.14_TL.a_Bg*). Leider wird nicht ausgeführt, wie diese intensivere akustische Repräsentation umgesetzt werden soll. Unabhängig von einem zusätzlichen Einsatz optischer Repräsentationen wird auch kritisiert, dass die Hörbeispiele zu kurz ausfallen (z. B. *Prof.1_TL.a_Bg*; *Prof.7_TL.a_Rf*). Trotz dieser Einschränkungen wird der Einsatz von akustischen Repräsentationen auch für diese Erklärung in einzelnen Aussagen positiv bewertet (z. B. *Hsl.11_TL.o_Rf*; *Stud.9_TL.o_Bg*). So trage das Vorspielen dazu bei, dass die Erklärung als weniger abstrakt wahrgenommen wird (*Hsl.5_TL.o_Bg*). Auch wird der Einsatz akustischer Repräsentationen als gut eingestuft, wenn die Adressat*innen über musikalische Vorerfahrung verfügen (*Lk.16_TL.o_Bg*).

Ein anderes Bild zeigt sich bei der Auswertung für die Erklärung *TA_a*. Hier wird kaum in Frage gestellt, dass Schüler*innen Unterschiede zwischen den Taktarten hörend unterscheiden können. Vielmehr wird dort der Verzicht auf optische Repräsentation positiv hervorgehoben, da damit eine Fokussierung auf das Hören einhergeht (z. B.

Hsl.8_TA.a_Bg; Lk.3_TA.a_Bg). Darüber hinaus werden in den Aussagen zu diesem Video inhaltlich-methodische Empfehlungen gegeben, die sich vor allem auf zwei Aspekte richten: Zum einen wird darauf hingewiesen, dass der Effekt der Verfremdung durch das Klatschen des Rhythmus‘ zum vermeintlich nicht passenden 3/4-Takt deutlicher herauskommen kann, wenn das Pattern zweimal geklatscht wird (*Prof.8_TA.a_Vb*) – das erinnert an die Empfehlung, akustische Repräsentationen nicht zu kurz ausfallen zu lassen (*Prof.1_TL.a_Bg; Prof.7_TL.a_Rf*). Zum anderen wird vorgeschlagen, die akustische Repräsentation der Taktarten vollständig und gut hörbar zu gestalten. Vielfach wird dabei gefordert, den Puls durch lautes Mitzählen (z. B. *Prof.4_TA.a_Bg; Lk.20_TA.a_Vb; Stud.41_TA.a_Vb*) hörbar zu machen oder die Schüler*innen zum eigenen Mitzählen anzuregen (*Stud.35_TA.a_Bg*).

Bei den Erklärungen, die beide Repräsentationen beinhalten (NW_oa; TL_oa; TA_oa), treten in den offenen Antworten neben den bisher genannten Aspekten noch weitere hervor: Beispielsweise wird das Klatschen des Rhythmus‘ als zu abstrakt empfunden und ergänzend dazu eine klangliche Demonstration am Klavier vorgeschlagen (*Prof.8_TA.oa_Vb*). Im Kontext der Erklärung zu Notenwerten wiederum wird der Einsatz des Klaviers kritisiert, da die Notendauer mit einem verklingenden Instrument nicht adäquat dargestellt werden kann (*Hsl.11_NW.oa_Vb*). Auch sollten nicht, wie in der Erklärung, nur einzelne Töne gespielt werden; vielmehr sollte Musik erklingen, welche die Kinder kennen (*Hsl.4_NW.oa_Vb*).

Analog zur Empfehlung einer zusammenfassenden Visualisierung der drei Taktarten (*Hsl.8_TA.oa_Bg; Slk.12_TA.oa_Bg*) wird für die Erklärung der Notenwerte vorgeschlagen, diese am Ende akustisch zusammenzufassen und nacheinander zu spielen (z. B. *Hsl.8_NW.oa_Rf; Hsl.9_NW.oa_Rf; Stud.27_NW.oa_Bg; Stud.62_NW.oa_Vb*). Bemängelt wurde bei dieser Erklärung auch, dass das Metrum schlecht hörbar war (*Lk.3_NW.oa_Rf*) und dieses über die einzelnen Notenwerte hinaus weitergeführt werden sollte (*Lk.20_NW.oa_Rf*).

Neben den genannten Aspekten, die sich konkret auf das Verstehen der Schüler*innen beziehen, finden sich in den offenen Antworten noch weitere Faktoren, die jedoch nur vereinzelt thematisiert werden. Gefordert wird beispielsweise ein mehrmaliges Abspielen der Hörbeispiele (z. B. *Slk.15_TL.a_Rf; Lk.19_TA.a_Rf; Stud.62_TL.oa_Rf*). Eine Optimierung akustischer Repräsentationen könnte zudem durch Variation in Tempo oder Tonhöhe (z. B. *Hsl.11_NW.oa_Vb; Slk.12_NW.oa_Bg*), zusätzliche Hörbeispiele wie weitere Rhythmen oder bekannte Lieder gelingen (z. B. *Stud.28_TL.oa_Rf; Prof.7_NW.oa_Vb; Stud.56_TA.oa_Bg*).

Zusammenfassend unterstützen auch diese Auswertungen der codierten Textstellen zum Einsatz akustischer Repräsentationen die Erkenntnisse der *mixed ANOVA* (siehe Kap. 5.3). In den Aussagen der Studierenden und Lehrenden werden Klangbeispiele insbesondere als klangliche Form von Veranschaulichung betrachtet. Sie ist notwendig, da Musik als klingendes Phänomen nur auf diese Weise dargestellt werden kann. Bei der Frage, ob akustische Repräsentationen für Schüler*innen zu einer besseren Vorstellung beitragen können, ist möglicherweise entscheidend, welche musikalischen Vorerfahrungen und Fähigkeiten vorliegen. Insbesondere können rein akustische Darstellungen für

Schüler*innen ohne ausreichende musikalische Vorerfahrungen überfordernd sein. Für diese Gruppe an Adressat*innen erscheinen optische Repräsentationen zwingend notwendig.

Hinsichtlich der verschiedenen Arten einer akustischen Darstellung (z. B. Vorsingen, Vorklatschen oder Vorspielen am Instrument) können Unterschiede herausgearbeitet werden: So bleibt das Vorklatschen eines Rhythmus⁴ für Schüler*innen auf einer abstrakten Ebene, die um eine Klangdemonstration durch ein Instrument ergänzt werden sollte. Diese wiederum wird im Kontext der hörenden Unterscheidung von Ganz- und Halbtonschritten für Schüler*innen als zu komplex erachtet. Der Einsatz akustischer Repräsentationen ist demnach am Grad der Abstraktion sowie der Komplexität des repräsentierten Inhalts auszurichten.

Analogien und fachübergreifende Vergleiche

Im folgenden Abschnitt werden Aussagen zum Einsatz von Analogien oder (fachübergreifenden) Vergleichen zusammengefasst, auf die in den Erklärungen mündlich hingewiesen wird. Nicht gemeint sind demnach Analogien, die Visualisierungen zugrunde liegen – die Abbildung von Treppenstufen kann beispielsweise als Analogie für das stufenweise Auf- und Absteigen einer Tonleiter dienen. Vergleiche zu anderen Fächern können auch Analogien beinhalten, anhand derer weitere Zugänge zu einem Thema möglich sind und diese somit die Qualität einer Erklärung erhöhen. Solche Aussagen finden sich in den offenen Antworten insbesondere bei den Erklärungen zum Thema Notenwerte und, in abgeschwächter Form, zum Thema Tonleiter.

In der Erklärung NW_o setzt die Lehrkraft zwei verschiedene Analogien ein: Zum einen verweist sie auf Ähnlichkeiten mit dem Bruchrechnen im Fach Mathematik. Hier werden Unterschiede zwischen den zwei Statusgruppen deutlich: Während Studierende hinsichtlich der Verhältnisse zwischen einzelnen Notenwerten den Bezug zum Bruchrechnen eher positiv bewerten (z. B. *Stud.41_NW.o_Bg*; *Stud.35_NW.oa_Bg*) und Einschränkungen beispielsweise nur in Form abfallender Motivation von Schüler*innen vermuten (*Stud.3_NW.o_Bg*), bewerten die Lehrenden den Vergleich eher negativ. Sie beschreiben die Erklärung NW_o als „sehr mathematisch und damit sehr abstrakt“ (*Slk.15_NW.o_Bg*), den Mathematik-Bezug als umständlich (*Lk.24_NW.o_Bg*) und die Erklärung als musikfern (*Prof.1_NW.o_Bg*).

Zum anderen zieht die Lehrkraft einen Vergleich mit dem Konzept einer Zeitstunde – dabei wird eine Analogie zwischen der Unterteilung von Notenwerten in Ganze, Halbe und Viertelnoten und der Aufteilung einer ganzen Zeitstunde in Viertel- oder halbe Stunden hergestellt. Auch hier fällt auf, dass die Studierenden den Einsatz der Analogie grundsätzlich positiv bewerten (z. B. *Stud.17_NW.o_Bg*; *Stud.26_NW.o_Bg*; *Stud.44_NW.o_Bg*), während die Gruppe der Lehrenden mehr differenziert. Auch sie bewerten den Einsatz der Analogie auf positiv (z. B. *Slk.4_NW.o_Bg*; *Slk.7_NW.o_Bg*; *Lk.9_NW.o_Bg*), da dieser der Veranschaulichung dient (z. B. *Lk.2_NW.o_Bg*). Auch wird herausgestellt, dass sich die Analogie eignet, da es sowohl bei Notenwerten als auch bei der Zeitstunde um die Dauer geht (*Prof.3_NW.o_Bg*; *Stud.5_NW.o_Bg*) und dadurch zudem der Alltag beziehungsweise die Lebenswelt der Schüler*innen berücksichtigt

werden (z. B. *Prof.5_NW.o_Bg*; *Hsl.2_NW.o_Bg*; *Lk.13_NW.o_Bg*; *Stud.41_NW.o_Bg*). Allerdings wird der Einsatz beispielsweise als *Umweg* bezeichnet und damit negativ bewertet (z. B. *Slk.14_NW.o_Vb*; *Lk.23_NW.o_Bg*). Bemängelt wird auch, dass die Analogie nicht vollständig passend ist, da eine Zeitangabe von einer Viertelstunde nicht immer als exakt 15 Minuten verstanden wird (*Slk.3_NW.o_Rf*). Auch funktioniert die weitere Unterteilung in Achtel- und Sechzehntelnoten nicht mehr, da man weder von einer Achtel- noch einer Sechzehntelstunde spricht (*Hsl.3_NW.o_Bg*). Vereinzelt werden dadurch auch Fehlvorstellungen auf Seiten der Schüler*innen vermutet (*Hsl.2_NW.o_Bg*), etwa dass eine Viertelnote eine Viertelstunde lang klingt (*Stud.56_NW.o_Bg*).

Bei den Erklärungen zur Tonleiter wurde ein Legobaustein als Analogie dafür verwendet, dass ein Lied in einer Tonart in der Regel auf eine Auswahl bestimmter Töne zurückgreift, die als Bau- beziehungsweise Tonmaterial verstanden werden sollen. Auch hier sind die Bewertungen divergent, diesmal jedoch über die Statusgruppen hinweg. Einerseits wird der Vergleich mit dem Baustein als anschaulich und hilfreich eingeschätzt (z. B. *Slk.4_TL.oa_Bg*; *Slk.10_TL.oa_Bg*; *Lk.18_TL.oa_Vb*; *Stud.60_TL.oa_Bg*). Andererseits wird auch kritisiert, dass der Begriff Baustein ungünstig ist, da dieser beispielsweise in Zusammenhang mit Motiven oder Themen verwendet werde (z. B. *Prof.3_TL.oa_Bg*; *Lk.16_TL.oa_Vb*).

Zusammenfassend erscheint der Einsatz von Analogien ähnlich kontrovers wie der Einsatz der Klaviatur – bei dieser stellen die einzelnen Töne in der Unterscheidung zwischen weißen und schwarzen Tasten auch eine Analogie für den Aufbau einer Dur-Tonleiter dar. Einer Analogie wird einerseits zugesprochen, den Erklärgegenstand besser zu veranschaulichen und entsprechend der Ausgestaltung auch mit dem Alltag und der Lebenswelt der Schüler*innen zu verknüpfen. Andererseits sei der Einsatz dahingehend zu überprüfen, inwieweit die Analogie tragfähig ist. Während das Konzept Stunde für Notenwerte bei der Ganzen, der Halben und der Viertelnote funktioniert, kann die weitere Übertragbarkeit für Achtel- und Sechzehntelnoten zu Fragen und Missverständnissen auf Seiten der Lernenden führen.

Beispiele

Abschließend werden Aussagen zur Verwendung und Eignung des Fußballrhythmus' als musikalisches Beispiel zusammengefasst. Vereinzelt wird positiv bewertet, dass das Beispiel bekannt ist (*Stud.3_TA.oa_Vb*) oder originell verwendet wurde (*Slk.5_TA.oa_Vb*). Es überwiegen jedoch die negativen Einschätzungen. Kritisiert wird insbesondere, dass sich bei der Verfremdung des Rhythmus', der original im 4/4-Takt steht, für den 2/4-Takt und 3/4-Takt Überbindungen ergeben (z. B. *Hsl.5_TA.oa_Vb*; *Hsl.6_TA.oa_Vb*; *Slk.10_TA.oa_Vb*). Da sich das Beispiel deshalb nicht zur Umformung eignet, wird es als schlecht und für Schüler*innen zu komplex bewertet (z. B. *Prof.4_TA.a_Vb*; *Prof.6_TA.oa_Vb*; *Hsl.6_TA.oa_Vb*). Im 3/4-Takt ergibt sich zudem eine Synkope, die zum einen die Betonungsordnung stört und zum anderen bei den Schüler*innen zu Verwirrung führen kann (z. B. *Hsl.14_TA.oa_Vb*; *Slk.9_TA.oa_Vb*; *Slk.11_TA.oa_Vb*; *Lk.15_TA.oa_Vb*; *Lk.23_TA.oa_Vb*). Dabei wird insbesondere die optische Darstellung des Rhythmus' als problematisch angesehen, weshalb Erklärung TA_a ohne optische

Repräsentation von einzelnen Personen bevorzugt wird (*Lk.20_TA.a_Bg*). Vorgeschlagen wird zudem, neben diesem verfremdeten Rhythmusbeispiel für die Darstellung des 3/4-Taktes ein zusätzliches Musikbeispiel wie einen Walzer (z. B. *Lk.16_TA.0a_Bg*; *Lk.25_TA.0a_Vb*; *Stud.49_TA.0a_Vb*) oder ein den Schüler*innen bekanntes Lied zu verwenden (z. B. *Stud.34_TA.0a_Bg*; *Hsl.4_TA.0a_Vb*).

Anhand dieses konkreten Falls kann zusammenfassend für den Einsatz von Beispielen abstrahiert werden, dass diese auch unter methodisch-didaktischen Überlegungen und Veränderungen in sich stimmig sein müssen. Dabei ist es zwar von Vorteil, wenn der Rhythmus aus der Lebenswelt der Schüler*innen stammt, jedoch erscheint dies im Lichte der Aussagen nachrangig gegenüber der weiteren Passung bei didaktisch begründeten Veränderungen. Diese bereits dargelegte Einschränkung könnte durch den Einsatz weiterer musikalischer Beispiele abgemildert werden.

In der Gesamtschau kann der Einsatz musikbezogener Repräsentationen als bedeutendes Qualitätsmerkmal aufgefasst werden. Dabei werden die optischen Repräsentationen insbesondere für Schüler*innen mit wenig musikalischer Vorerfahrung als essenziell notwendig angesehen, während die akustischen zwar als ähnlich relevant, aber für dieselbe Gruppe an Lernenden als nicht hinreichend erachtet werden. Die Verwendung von Analogien, Vergleichen und musikalischen Beispielen dient der Veranschaulichung des Erklärgegenstandes. Bei ihrem Einsatz – egal ob in Form einer Visualisierung oder als mündlicher Hinweis – müssen jedoch Einschränkungen aufgrund didaktischer Entscheidungen reflektiert und in der Ausgestaltung berücksichtigt werden.

5.4.3 Subjektive Theorien zu weiteren Merkmalen

Neben den fünf ausgewählten Qualitätsmerkmalen, die a priori als deduktive Kategorien für die Auswertung festgelegt wurden, konnten im Analyseprozess noch weitere induktive Kategorien herausgearbeitet werden. Sie fassen Aussagen zu potenziellen Kandidaten weiterer Qualitätsmerkmale unterrichtlichen Erklärens zusammen. Folgende Aspekte konnten identifiziert werden: Neben Aussagen zum Professionskönnen der Lehrkraft thematisieren die Studienteilnehmer*innen den Musikbezug in der Erklärung, die fachlich korrekte Darstellung der Gegenstände sowie die Menge an Informationen. Während zu all diesen Aspekten ein klarer musikspezifischer Bezug in den Aussagen deutlich wird, weisen die Textstellen zum inhaltlichen Tempo der Erklärung eher allgemeine Aussagen auf, die in ihrem Wortlaut zum Großteil auch für Erklärungen anderer Fächer bestehen bleiben könnten.

Das Professionskönnen der Lehrkraft spielt in den Aussagen eine unbedeutende Rolle. Während beispielsweise das kurze Vorspiel des Liedes *Dracula Rock* zum Einstieg von Erklärung TL_a als „zu brav, langweilig und wenig mitreißend“ (*Hsl.1_TL.a_Rf*) beschrieben wurde und „etwas pffiffiger“ (*Hsl.8_TL.a_Rf*) hätte ausfallen können, wurde das gleichmäßige Spiel in Erklärung TL_0a gelobt (*Lk.19_TL.0a_Rf*). Neben solchen Einzelaussagen wird etwas häufiger angesprochen, dass das Klatschbeispiel der Erklärungen zum Thema Taktarten nicht ganz flüssig vorgetragen wurde (z. B. *Stud.58_TA.0a_Bg*; *Stud.3_TA.a_Rf*). Aus diesen Einzelaussagen kann jedoch keine Systematik

herausgearbeitet werden, die Hinweise zu einem Qualitätsmerkmal unterrichtlichen Erklärens geben. Damit reiht sich das Professionskönnen in die anderen, eher unbedeutenderen, Kategorien wie die sprecherischen und körpersprachlichen Aspekte aus dem Bereich der Performanz der Lehrkraft ein. Ebenfalls in Form von Einzelmeinungen werden an wenigen Stellen im Textkorpus ein „fehlender Bezug zur Musik“ (*Slk.15_NW.o_Bg*) im Allgemeinen oder solcher „Musik, wie sie die Kinder täglich erleben“ (*Hsl.4_TA.a_Bg*) bemängelt oder die Erklärung NW_oa beispielsweise als „musikfern“ (*Prof.1_NW.oa_Bg*) bewertet.

Eine bedeutendere Rolle in den Aussagen der Studierenden und Lehrenden spielt die fachliche Korrektheit der Erklärungen – und zwar über alle drei Themen annähernd gleich verteilt. Dabei überrascht, dass für dieselben Erklärungen unterschiedliche und teils gegensätzliche Einschätzungen über einzelne Fachinhalte vorliegen – auch innerhalb einzelner Teilstatusgruppen wie (Seminar-)Lehrkräften, von denen ein einheitliches Verständnis über die fachliche Korrektheit musikkundlicher Inhalte erwartet wurde.

Den Erklärungen zum Thema Notenwerte wird einerseits fachliche Korrektheit attestiert (z. B. *Hsl.11_NW.oa_Rf*; *Slk.10_NW.oa_Bg*; *Slk.11_NW.o_Bg*), andererseits identifizieren die Studienteilnehmer*innen „fachliche Unschärfen“ (*Slk.14_NW.oa_Bg*). Als verkürzt und damit falsch wird die Gleichstellung von Schlag und Viertelnote bezeichnet (z. B. *Slk.7_NW.oa_Bg*; *Slk.9_NW.o_Bg*; *Lk.22_NW.o_Bg*), wenn nicht explizit darauf hingewiesen wird, dass ein Schlag in einer anderen Taktart beispielsweise auch einer halben Note entsprechen kann (*Stud.22_NW.o_Bg*). Auch die Aussage, dass die Orientierung an Schlägen präziser sei, wird als fachlich falsch beziehungsweise unsinnig bewertet (z. B. *Hsl.4_NW.o_Vb*; *Lk.22_NW.o_Bg*). In diesem Kontext wird generell angemerkt, dass Schlag weniger mit Notenwerten als mit Takt zu tun hat (*Slk.5_NW.o_Vb*).

Bei den Erklärungen zum Thema Taktarten wird die fachliche Korrektheit der Erklärung ebenfalls eher angezweifelt (z. B. *Prof.6_TA.oa_Bg*). So wird bemängelt, dass „die Begriffe Metrum (Lehre von der Betonung) und Taktart (Gruppierung einer bestimmten Anzahl von Grundschlägen) gleichgesetzt“ werden (*Lk.15_TA.oa_Bg*; z. B. auch *Prof.1_TA.oa_Bg*). Dadurch ergibt sich eine Hierarchie, in welcher „der Rhythmus dem rein strukturgebenden Takt untergeordnet wird“ (*Slk.14_TA.oa_Vb*). Fachlich korrekter, da auch musikalisch-ästhetisch wertvoller, wäre demnach, Takt und Puls dem Rhythmus unterzuordnen (*Slk.14_TA.a_Vb*).

Im Gegensatz zu diesen beiden Themen werden die Erklärungen zur Tonleiter überwiegend als fachlich korrekt bewertet (z. B. *Slk.5_TL.a_Bg*; *Prof.8_TL.oa_Bg*; *Hsl.11_TL.o_Rf*). Einzig die Aussage, dass Töne eines Liedes stufenweise angeordnet eine Tonleiter ergeben, wurde ohne weitere Differenzierung als verkürzt bewertet. Zum einen können leitereigene Töne in einem Lied fehlen, zum anderen auch leiterfremde Töne vorkommen; weiterhin ist denkbar, dass der Ambitus eines Stücks über eine Oktave hinausgeht (z. B. *Hsl.12_TL.o_Bg*; *Lk.18_TL.o_Bg*; *Lk.2_TL.o_Bg*; *Stud.26_TL.o_Bg*).

Ein weiterer, inhaltlicher Aspekt der Erklärung stellt die Menge der vermittelten Informationen dar. Tendenziell kann insbesondere bei den Erklärungen TA_a und TL_a (Erklärungen ohne optische Repräsentation) die Kritik beobachtet werden, dass zu viele Informationen in der Erklärung gegeben wurden (z. B. *Hsl.8_TL.a_Rf*; *Lk.5_TL.a_Bg*;

Stud.8_TL.a_Bg; Hsl.11_TA.a_Vb; Slk.15_TA.a_Rf; Lk.15_TA.a_Bg). Obwohl auch in den Aussagen über andere Erklärungen ein Überfluss an Informationen bemängelt wird (z. B. *Slk.10_TA.aa_Bg; Lk.3_TA.aa_Rf; Stud.54_NW.aa_Bg; Slk.6_TL.aa_Bg; Lk.24_NW.o_Rf*), könnte ein Zusammenhang zwischen zu viel Informationen und dem Erklärmodus bestehen, der ausschließlich eine akustische und keine optische Repräsentation beinhaltet. Das Fehlen notwendiger Informationen wird hingegen auf Inhalte bezogen. Im Kontext des Themas Tonleiter wurde bemängelt, dass weder Begriffe wie *Grundton* oder *Oktave* noch die Anzahl und Namen der Töne in einer Tonleiter erklärt wurden (z. B. *Slk.15_TL.a_Bg; Stud.9_TL.a_Bg; Slk.5_TL.o_Vb; Lk.15_TL.aa_Bg*). Bei den Erklärungen zu Taktarten sollten sowohl die Nebenbetonungen sowie die Anzahl möglicher Schläge pro Taktart erklärt werden (z. B. *Stud.39_TA.aa_Bg; Stud.61_TA.aa_Bg*). Auch wurde vorgeschlagen, die Erklärung mit Informationen durch den Einbezug von Vorstellungshilfen wie *Walzertakt* oder *Marschtakt* anzureichern und unterschiedliche Wirkungen der Taktarten zu beschreiben (z. B. *Slk.10_TA.aa_Bg; Slk.10_TA.a_Bg*).

Neben diesen musikspezifischen Qualitätsmerkmalen bezieht sich der letzte Aspekt auf das inhaltliche Tempo der Erklärung. Gemeint ist damit, wie schnell die Erklärung zu neuen Denkschritten weitergeht oder ob beispielsweise Wiederholungsschleifen eingebaut sind. Dabei wurde äußerst selten angemerkt, dass das inhaltliche Tempo passend ist; keine Aussagen liegen dazu vor, dass die Erklärung an einzelnen Stellen zu langsam fortgeschritten ist; hingegen wird das Tempo in den allermeisten Aussagen als zu schnell bewertet (z. B. *Hsl.2_TL.o_Bg; Stud.32_TL.o_Bg; Lk.15_NW.aa_Bg*). Dabei tragen die meisten Aussagen nicht zur inhaltlichen Ausschärfung des Aspekts bei: In knappen Bemerkungen wurde die Erklärung als „zu schnell“ (*Lk.17_TA.aa_Bg; Lk.23_TA.aa_Rf*) „etwas zu schnell“ (*Lk.11_TA.a_Bg*) oder „viel zu schnell“ (*Slk.12_TA.a_Bg*) beurteilt oder sollte „langsamer“ (*Hsl.11_TA.a_Vb*) ablaufen. Konkrete Aussagen darüber, welcher Denkschritt genau zu schnell war, liegen hingegen nur sehr vereinzelt vor: Bemängelt wurde beispielsweise, dass der Übergang von Halbtonschritten zu Ganztonschritten sehr schnell und für Schüler*innen womöglich nicht nachzuvollziehen war (z. B. *Hsl.14_TL.a_Bg; Stud.15_TL.a_Bg*). Moniert wurde auch das zu schnelle Herausfiltern der Töne aus dem Lied *Dracula Rock* (z. B. *Stud.35_TL.aa_Bg; Slk.10_TL.aa_Bg*).

Zusammenfassend können zwei weitere potenzielle Kandidaten für Qualitätsmerkmale unterrichtlichen Erklärens im Fach Musik in den Daten identifiziert werden: Die fachliche Korrektheit spielt ebenso eine Rolle wie eine angemessene Menge an Informationen. Weiterhin kann auch das inhaltliche Tempo der Erklärung als potenzielles Qualitätsmerkmal verstanden werden. Andere Aspekte wie Musikbezug der Erklärung oder das Professionskönnen der Lehrkraft können auf Basis der Daten noch nicht zu möglichen Qualitätsmerkmalen gezählt werden.¹⁴⁹

149 Alle weiteren Kategorien und Subkategorien des Kategoriensystems (siehe *Kategoriensystem 2b, Anhang F4*) sind in der vorliegenden Darstellung nicht berücksichtigt, da sie für die Frage nach Qualitätsmerkmalen unterrichtlichen Erklärens nicht hinreichend erschienen.

5.4.4 Zusammenfassung

Hinsichtlich der Frage, welche Merkmale für die Qualität unterrichtlichen Erklärens im Fach Musik bedeutsam sind, muss die Antwort verschiedene Dimensionen berücksichtigen. Die Exploration der Merkmale erfolgte zum einen im Kontext des Erklärens musiktheoretischer Unterrichtsinhalte.¹⁵⁰ Demnach erhebt die Auflistung keinen Anspruch auf Vollständigkeit sowie auf Abbildung aller Bereiche des Musikunterrichts. Die Darstellung folgt der Struktur des Abschnitts 5.4: Zunächst wird eine Zusammenfassung der fünf in der Konzeptualisierung ausgewählten und den Regressionsanalysen untersuchten Merkmale unter Berücksichtigung der Zugehörigkeit zur Statusgruppe vorgenommen. Anschließend erfolgt eine Ausdifferenzierung der Merkmale unter Berücksichtigung einzelner Facetten sowie die Ergänzung um potenziell weitere Kandidaten. Bei diesem zweiten Schritt werden Einschränkungen in der Reichweite der Aussagen berücksichtigt.

Die fünf ausgewählten Merkmale unterrichtlichen Erklärens erscheinen aus verschiedenen Gründen als bedeutsam: Zum einen kann für alle anhand eines über alle Studienteilnehmer*innen hinweg gerechneten gemischten Modells ein signifikanter Einfluss nachgewiesen werden (siehe Tab. 5.41), wobei eine Aufteilung nach Statusgruppen zu differenzierten Ergebnissen führt. Dabei bleibt ein signifikanter Einfluss für jedes der fünf Merkmale bei mindestens einer Statusgruppe bestehen. Aufgrund der Tatsache, dass keines der Merkmale eine statistisch signifikante Bedeutung in allen Statusgruppen besitzt, muss von einer perspektivenspezifischen Validität der Qualitätsmerkmale ausgegangen werden.

Auf Seiten der Schüler*innen zeigen sich dabei insbesondere die eher auf der Oberflächenebene verorteten Merkmale *Sprech- und Körperausdruck* sowie der *Einsatz musikbezogener Repräsentationen* einen signifikanten Einfluss auf das Globalurteil – der ebenfalls signifikante Effekt des Merkmals *Sprachliche Verständlichkeit* muss dagegen aufgrund niedriger Skalenreliabilitätswerte in Frage gestellt werden (siehe Tab. 5.37).

Bei den Studierenden hingegen liegen signifikante Effekte für die Merkmale *Strukturiertheit* und *Adressatenorientierung* vor (siehe Tab. 5.38). Der negative Einfluss des Merkmals *Sprachliche Verständlichkeit* führt auf den ersten Blick zu einem kontraintuitiven Ergebnis, könnte aber unter anderem auf den Professionalisierungsprozess zurückzuführen sein, in dem sich Studierende während der Ausbildung befinden. Aufgrund der Thematisierung in den offenen Antworten wird das Merkmal dennoch als relevant eingestuft, wengleich die Operationalisierung von *Sprachliche Verständlichkeit* Ziel weiterer Untersuchungen sein sollte.

Auf Seiten der Lehrenden kann ausschließlich für das Merkmal *Adressatenorientierung* eine statistisch signifikante Bedeutung nachgewiesen werden (siehe Tab. 5.37). Es liegt die Vermutung nahe, dass die Gruppe der Lehrenden *Adressatenorientierung* als Leitprinzip versteht, unter dessen Bewertung die anderen Merkmale teilweise subsumiert werden.

¹⁵⁰ In weiteren empirischen Forschungsprojekten wäre zu untersuchen, ob die Qualitätsmerkmale auch für andere Bereiche des Musikunterrichts Gültigkeit haben.

Da in der vorliegenden Arbeit die unterschiedlichen Sichtweisen auf unterrichtliches Erklären anhand der verschiedenen Statusgruppen gleichberechtigt behandelt werden und sich gegenseitig ergänzen sollen, kann für unterrichtliches Erklären Folgendes zusammengefasst werden: Gutes Erklären im Musikunterricht bedarf

- einer Orientierung an Adressat*innen,
- eines strukturierten Ablaufs,
- einer sprachlichen Verständlichkeit,
- des Einsatzes sowohl optischer als auch akustischer Repräsentationen sowie
- eines angemessenen Sprech- und Körperausdrucks der erklärenden Lehrkraft.

Weitere Differenzierungen zu diesen einzelnen Merkmalen liefern die offenen Antworten, die allerdings nur von Studierenden und Lehrenden abgegeben wurden. Aufgrund der Aussagen, mit denen die Bewertung der Erklärqualität begründet wurde, können für diese Gruppen folgende der genannten Merkmale als bedeutsam für die Qualität unterrichtlichen Erklärens eingeschätzt werden: Insbesondere zu *Adressatenorientierung*, *Strukturiertheit* und *Einsatz musikbezogener Repräsentationen*, aber auch zu sprachlichen Aspekten konnten viele Aussagen codiert werden, die von den Studienteilnehmer*innen erhoben wurden, bevor diese die kriterienbezogenen Items im zweiten und dritten Durchgang des Fragebogens beantwortet haben. Der *Sprech- und Körperausdruck* der Lehrkraft wurde von beiden Gruppen kaum thematisiert und wird demnach für Studierende und Lehrende als nicht relevant erachtet. Diese Beobachtungen stützen weitestgehend die Ergebnisse der Regressionsmodelle.

Beide Statusgruppen nennen als Facetten des Merkmals *Adressatenorientierung* insbesondere die Orientierung am Vorwissen und an der Lebenswelt der Schüler*innen sowie die Berücksichtigung der Heterogenität in Bezug auf musikalische Vorerfahrungen. Während in den Regressionsmodellen für das Merkmal *Strukturiertheit* nur bei den Studierenden ein signifikanter Effekt nachgewiesen werden kann, rücken es beide Gruppen in den offenen Antworten gleichermaßen in den Vordergrund. Drei Facetten werden insbesondere deutlich: Für den Einstieg in eine Erklärung wird das Hören oder Singen von Musik herausgestellt. Ohne dass in diesem Zusammenhang ein konkreter Zweck deutlich wird, könnten beispielsweise motivationale Ziele vermutet werden.

Deutlich wurde zweites, dass in den einzelnen Erklärschritten zunächst Prinzipien und Konzepte erklärt und dargestellt werden sollten, bevor eine akustische Demonstration oder eine musizierende Auseinandersetzung erfolgen sollte. Drittens wurde mehrfach eine Zusammenfassung am Ende der Erklärung gefordert. Diese kann je nach Thema mit einer Visualisierung oder einer akustischen Demonstration umgesetzt werden; denkbar sind dabei gleichermaßen optische und akustische Repräsentationen. Die *Sprachliche Verständlichkeit* einer Erklärung kann auf Basis der Aussagen der Studienteilnehmer*innen vor allem anhand von zwei Aspekten qualitativ hochwertig umgesetzt werden: Zum einen sind Formulierungen wie „Puls oder Takt darunter denken“ (*Lk.15_TA.a_Bg*) auf ihre (Miss-)Verständlichkeit hin zu überprüfen. Zum anderen sollten insbesondere Fachbegriffe nur dann verwendet werden, wenn sie ausreichend erklärt werden oder für

Schüler*innen bereits klar sind. Hinsichtlich *Sprech- und Körperausdruck* kann anhand der offenen Antworten nur das Sprechtempo der Lehrkraft als gering bedeutsam eingestuft werden. Andere Aspekte der Operationalisierung wurden teilweise nur einmal oder gar nicht genannt, zusätzliche Facetten konnten in den offenen Antworten nicht beobachtet werden.

Der *Einsatz musikbezogener Repräsentationen* stellt ein weiteres bedeutsames Merkmal unterrichtlichen Erklärens im Fach Musik dar. Unterschieden werden muss dabei zwischen verschiedenen Formen. Die Aussagen der Studienteilnehmer*innen legen nahe, dass optische Repräsentationen für das Verstehen von Schüler*innen wichtig sind – insbesondere, wenn diese über keine musikalische Vorerfahrung verfügen. Sinnvoll ist die Abbildung des Notentextes dabei nicht nur zur Verifikation des Erklärten, sie dient auch der Veranschaulichung musikalischer Phänomene. Markierungen im Notentext können zudem dabei helfen, relevante Informationen herauszuheben. Eine Gegenüberstellung verschiedener Notenbeispiele kann für einen Überblick am Ende der Erklärung hilfreich sein. Andere Formen der Visualisierung, wie eine pyramidale Anordnung von Notenwerten oder eine Klaviatur für die Abfolge von Halb- und Ganztonschritten in einer Tonleiter, können der Erklärung als Hilfestellung dienen. Limitationen, wie die verkürzte Darstellung der Klaviatur (z. B. ergeben die weißen und schwarzen Tasten nur ein sinnvolles Abbild für die Abfolge von Ganz- und Halbtonschritten für die C-Dur-Tonleiter), sollten dabei in der Erklärung immer angesprochen und explizit deutlich gemacht werden.

Akustische Repräsentationen werden ebenfalls als relevante Form erachtet. Sie sollten insbesondere für Schüler*innen mit fehlender musikalischer Vorerfahrung nicht für sich alleinstehen, sondern in Kombination mit einer optischen Darstellung erfolgen. Im Gegensatz zu optischen Repräsentationen, die für das Verstehen als zwingend und notwendig erachtet werden, wird akustischen Repräsentationen eher die Funktion einer Ergänzung zugewiesen. Sie sind zwar für ein umfassendes und klangliches Veranschaulichen musikalischer Phänomene notwendig, als alleinige Form der Darstellung jedoch für kognitive musikalische Ziele nicht hinreichend. Für klangliche Demonstrationen sollte ein passendes Instrument gewählt werden. Auch ist zu berücksichtigen, inwieweit gerade Schüler*innen ohne musikalische Vorerfahrung einer solchen klanglichen Demonstration folgen können, ohne überfordert zu sein. Andere Formen akustischer Repräsentation, wie das Vorklatschen eines Rhythmus', sollten auf ihren Grad an Abstraktheit überprüft und gegebenenfalls durch klingende Musik ergänzt werden.

Der Einsatz von Analogien wird kontrovers diskutiert. Einerseits werden Konzepte wie das der Zeitstunde als Analogie zur Aufteilung von Notenwerten als hilfreich bewertet, da sie aus der Lebenswelt der Schüler*innen stammen. Andererseits sind solche Analogien nur bedingt geeignet, wenn beispielsweise weitere Unterteilungen in Achtel- und Sechzehntelnoten nicht mehr passend abgebildet werden können. Gleiches gilt für die Klaviatur als Hilfsmittel und Analogie für die Anordnung von Ganz- und Halbtonschritten einer Dur-Tonleiter. Deutlich werden in diesem Zusammenhang persönliche Vorlieben und Erfahrungen der Studienteilnehmer*innen, die eine eindeutige Empfehlung für oder gegen den Einsatz einer solchen Analogie oder eines Vergleichs als Facette dieses Merkmals unterrichtlichen Erklärens erschweren. In jedem Fall sollten Limitationen und

Einschränkungen in der Nutzung einer Analogie sichtbar gemacht werden. Für den Einsatz von Musikbeispielen wird empfohlen, mehrere Beispiele einzusetzen, um ein musikalisches Phänomen exemplarisch abzubilden.

Ergänzt werden kann diese Differenzierung der Merkmale um potenziell weitere Kandidaten für Qualitätsmerkmale unterrichtlichen Erklärens. Diese sind fachliche Korrektheit, die passende Menge an Informationen sowie ein angemessenes inhaltliches Tempo der Erklärung. In den Darstellungen werden Wechselwirkungen zwischen den einzelnen Merkmalen deutlich. So richtet sich beispielsweise der Einsatz von optischen und akustischen Repräsentationen vor allem an die musikalischen Vorerfahrungen der Adressat*innen. Zudem können Visualisierungen oder akustische Darstellungen strukturierende Funktion übernehmen und darüber hinaus sowohl zu Beginn einer Erklärung als auch am Ende in Form einer Zusammenfassung zum Einsatz kommen.

Inhaltliche Aspekte, wie eine angemessene Menge an Informationen, hängen stark mit der Facette von *Strukturiertheit* zusammen, sich auf das Wesentliche zu konzentrieren. In diesem Kontext wird insbesondere Erklärungen mit rein akustischer Repräsentation ein Überfluss an Informationen attestiert, sodass ein Zusammenhang zwischen der Menge an Informationen und dem Erklärmodus vermutet werden kann. Diese Wechselwirkungen empirisch zu überprüfen, kann ein lohnendes Ziel zukünftiger Forschungsvorhaben sein.

Die oben genannten fünf Merkmale können auf Basis dieser Befunde weiter ausdifferenziert werden. Gutes Erklären im Musikunterricht bedarf demnach

- einer Orientierung am Vorwissen der Schüler*innen,
- einer Orientierung an der Lebenswelt der Schüler*innen,
- einer Berücksichtigung heterogener musikalischer Vorerfahrungen der Schüler*innen,
- eines musikalischen Einstiegs (singend, musizierend oder hörend),
- zunächst der Darstellung von Prinzipien oder Konzepten, bevor eine akustische Umsetzung erfolgt,
- einer Zusammenfassung in Form optischer beziehungsweise akustischer Repräsentation am Ende,
- der bewussten Verwendung von bekannten oder ausreichend erklärten Fachbegriffen,
- einer sorgsamem Verwendung von (Fach-)Formulierungen,
- eines angemessenen Sprechtempos der erklärenden Person,
- einer optischen Darstellung von Notentext,
- des angemessenen Einsatzes von Visualisierungen als Hilfsmittel,
- einer klanglichen Darstellung, die jedoch gerade für leistungsschwache Schüler*innen nur in Verbindung mit einer optischen Darstellung erfolgen sollte,
- verschiedener Formen der klanglichen Darstellung (z. B. Vorklatschen und Vorspielen am Instrument),
- der Auswahl passender Instrumente, um musikalische Phänomene darzustellen,

- einer Klärung potenzieller Limitationen bei der Verwendung von Analogien sowie
- des Einsatzes passender Beispiele.

Weitere potenzielle Kandidaten für gutes Erklären im Musikunterricht sind außerdem

- die fachliche Korrektheit einer Erklärung,
- die passende Menge an Informationen sowie
- ein angemessenes inhaltliches Erklärtempo.

5.5 Zusammenhang zwischen Vorwissen der Schüler*innen und Erklärqualität

F7) Wie hängt das Vorwissen der Schüler*innen mit deren Bewertungen zur Erklärqualität zusammen?

Dieser Abschnitt beantwortet Forschungsfrage sieben (F7), wie das Vorwissen der Schüler*innen mit deren Bewertungen zur Erklärqualität zusammenhängt. Aufgrund empirischer Erkenntnisse aus anderen Fächern beziehungsweise anderen Lehr-/Lernkontexten wird erwartet, dass die Höhe des Vorwissens signifikant bedeutsam für die Bewertungen der Erklärqualität ist (*Hypothese 7.1*).

Für eine nachvollziehbare Darstellung werden zunächst die deskriptiven Ergebnisse des Vorwissenstests pro Erklärthema präsentiert, bevor anschließend ein gemischtes lineares Regressionsmodell berechnet wird, das die Personen- und Videostuktur als *random effects* berücksichtigt, und das Globalurteil zur Erklärqualität als abhängige Variable durch zwei Prädiktoren vorhersagen lässt. Zum einen wird der Vorwissensscore als Gesamtscore über alle drei Erklärthemen als *fixed effect* berücksichtigt. Da in der *mixed ANOVA* (siehe Kap. 5.3.3, Tab. 5.25) für die Statusgruppe der Schüler*innen auch der Erklärmodus als signifikanter Effekt für die Bewertung der Erklärqualität der Schüler*innen nachgewiesen werden konnte, wird dieser zum anderen als zweiter Prädiktor ins Modell integriert.

Deskriptive Ergebnisse

Der Vorwissenstest besteht aus drei Teilen, bei denen pro Erklärthema drei Aufgaben gestellt wurden. Für jede richtige Antwort gab es einen Punkt, sodass pro Thema jeweils maximal drei Punkte möglich waren. Für den Gesamtscore ergibt sich dadurch eine mögliche Höchstpunktzahl von neun Punkten. Die Aufteilung der jeweiligen Scores pro Thema kann Tabelle 5.41 entnommen werden. Dabei zeigt sich einerseits, dass das größte Vorwissen bei den Schüler*innen zum Thema Notenwerte vorlag. Dies wurde bereits als möglicher Grund dafür vermutet, dass die Schüler*innen hier beide Erklärungen im

Schnitt mit den besten Globalurteilen bewertet haben (siehe Kap. 5.3.1, Tab. 5.14). Die Scores für Taktarten und Tonleiter fallen im Schnitt deutlich niedriger aus und liegen nahe beieinander. Dabei kann für das Thema Tonleiter im Schnitt ein etwas höheres Vorwissen als für das Thema Taktarten angenommen werden.

Tabelle 5.41 Anzahl der unterschiedlichen Scores sowie durchschnittlicher Score pro Erklärthema in den **Vorwissenstests** für alle Schüler*innen ($N = 142$)

Erklärthema	Notenwerte	Taktarten	Tonleiter
Score	n	n	n
0/3	3	57	34
1/3	24	46	57
2/3	40	27	34
3/3	75	12	17
M (Score)	2,32	0,96	1,24

Bem.: M : Mittelwert; n : Anzahl der Scores zu den Vorwissenstests getrennt nach Themen.

Für die nachfolgende Regressionsanalyse fließt der Gesamtscore über alle drei Vorwissenstests als gemeinsamer Prädiktor in das Modell ein. Dazu liegen Ergebnisse auf einer zehnstufigen Skala (Minimum: 0 Punkte; Maximum: 9 Punkte) für insgesamt $N = 142$ Schüler*innen vor. Die Darstellung der Verteilung dieser Werte dient dazu, einen Überblick über die Variable zu erhalten. Tabelle 5.42 zeigt die Häufigkeitswerte der einzeln erreichten Scores.

Tabelle 5.42 Verteilung der Gesamtscores aller Schüler*innen ($N_{max} = 142$)

Gesamtscore (Anzahl Punkte)	0	1	2	3	4	5	6	7	8	9
Häufigkeit	1	3	18	25	25	29	17	14	8	2
M (SD)	4,51 (1,89)									

Bem.: M : Mittelwert; SD : Standardabweichung.

Im Schnitt erreichen die Schüler*innen 4,51 Punkte, die Standardabweichung des Gesamtscores beträgt 1,89. Der Median liegt bei dem Score von 4 Punkten, der Modus bei 5 Punkten. Zwei Schüler*innen erreichen volle Punktzahl. Die Verteilung der Noten folgt keiner Normalverteilung (Überprüfung mittels *Kolmogorov-Smirnov*-Test). Nach Bortz und Schuster (2010, S. 214) kann aufgrund der Stichprobengröße dennoch von robusten Ergebnissen ausgegangen werden.

Linear Mixed Model

Neben der Variable Vorwissen, die für die nachfolgende Analyse z -standardisiert wird und als Prädiktor in das Regressionsmodell einfließt, fungiert der Erklärmodus als zweiter

Prädiktor, da dessen signifikanter Einfluss auf die Bewertung der Erklärqualität der Schüler*innen bereits in Kapitel 5.3.3 gezeigt werden konnte. Aufgrund der für das Modell notwendigen Dummy-Codierung (siehe Hilbert et al., 2019) wird der Faktor Erklärmodus aufgespalten, sodass einerseits der Einfluss für eine fehlende optische Repräsentation und andererseits für die fehlende akustische Repräsentation angegeben werden kann. Somit fließen insgesamt drei Prädiktoren als *fixed effects* in das Modell ein, während die Personen- und Videostruktur als *random effects* berücksichtigt werden. Tabelle 5.43 zeigt die Ergebnisse des hierarchischen linearen Regressionsmodells für die Schüler*innen mit dem Kriterium Globalurteil und den Prädiktoren Erklärmodus und Vorwissen sowie deren Interaktion.

Tabelle 5.43 Hierarchisches lineares Regressionsmodell mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach **Schüler*innen** ($N_{max} = 142$) und den sieben Videos geordneten Datenstruktur

Schüler*innen ($N_{Beob.} = 994$) ¹⁵¹		
ICC	%	38,81
Konstante	<i>b</i>	2,05**
Vorwissen.z	<i>b</i>	-0,25**
Erklärmodus.1 (<i>akustisch fehlt</i>)	<i>b</i>	0,05
Erklärmodus.2 (<i>optisch fehlt</i>)	<i>b</i>	0,51*
Vorwissen.z x Erklärmodus.1	<i>b</i>	0,05
Vorwissen.z x Erklärmodus.2	<i>b</i>	0,13*
R^2 (marginal konditional)		0,08 0,37

Bem.: *N*: Personenanzahl; $N_{Beob.}$: Anzahl der Beobachtungen ($N \times 7$ Videos); ICC: Intraklassenkorrelation; *b*: Regressionskoeffizient; R^2 : Determinationskoeffizient; *: $p \leq 0,05$; **: $p \leq 0,01$.

In diesem Modell¹⁵² besitzen die Prädiktoren *Vorwissen.z*¹⁵³ und *Erklärmodus.2* sowie die Interaktion zwischen *Vorwissen.z* und *Erklärmodus.2* signifikante Einflüsse auf die Globalbewertung: Die Höhe des Vorwissens der Schüler*innen hat einen statistisch signifikanten Einfluss auf die Bewertung der Erklärqualität. Das Regressionsgewicht von -0,25 bedeutet, dass in der Referenzkategorie das Globalurteil im Schnitt um eine Viertelnote niedriger und damit besser bewertet wird, wenn das Vorwissen im Schnitt um einen Punkt steigt. Schüler*innen bewerten demnach Erklärungen im Schnitt besser, je mehr sie vorher schon wissen. Das ist einerseits ein weiteres Indiz für die Vermutung,

151 Die Anzahl an Beobachtungen ist nun höher als im vorherigen Modell für die Gruppe der Schüler*innen (siehe Tab. 5.36), da alle Schüler*innen in die Regressionsanalyse einfließen und diesmal nicht die Zahl der Beobachtungen aufgrund der fehlenden Werte für das Merkmal *Sprech- und Körperausdruck* eingeschränkt sind.

152 Die Interklassenkorrelation beträgt 38,81 %.

153 *Vorwissen.z* steht für die z-standardisierte Variable Vorwissen. Die Standardisierung wurde vorgenommen, da so das Modell besser interpretierbar ist. Die Konstante kann demnach als Wert für durchschnittliches Vorwissen und den Erklärmodus interpretiert werden, der weder auf akustische noch auf optische Repräsentation verzichtet und für die Kombination aus beiden Repräsentationsformen steht.

dass das hohe Vorwissen zum Teil erklären kann, warum die Schüler*innen beide Erklärungen zum Thema Notenwerte am besten bewerten, während die Bewertung der Studierenden und Lehrenden durch den Erklärmodus und unabhängig vom Thema vorhergesagt werden kann. Für die Bewertung der Erklärqualität durch die Schüler*innen bedeutet es andererseits auch, dass die Objektivität der Bewertung eingeschränkt ist.

Die Regressionskoeffizienten des zweiten Prädiktors Erklärmodus, der aufgrund der Dummy-Codierung aufgespalten ist, unterstützen die Befunde aus der *mixed ANOVA* (siehe Kap. 5.3.3): Aufgrund der *z*-Standardisierung der Variable Vorwissen und der Dummy-Codierung für die Variable Erklärmodus entspricht die Konstante (2,05) der durchschnittlichen Globalnote für den Erklärmodus *optisch und akustisch*, da weder das Vorwissen noch die beiden Erklärmodi berücksichtigt werden. Im Verhältnis dazu fällt für die Bewertung des Globalurteils auch in diesem Modell statistisch nicht bedeutsam ins Gewicht ($b = 0,05$), wenn eine akustische Repräsentation fehlt (*Erklärmodus.1*). Dem Fehlen einer optischen Repräsentation kann jedoch eine statistisch signifikante Bedeutung nachgewiesen werden. Der Regressionskoeffizient von $b = 0,51$ bedeutet, dass das Globalurteil im Schnitt um 0,51 Notenstufen schlechter ausfällt, wenn eine optische Repräsentation fehlt.

Der Interaktionseffekt zwischen *Vorwissen.z* und *Erklärmodus.2* hat einen signifikanten Regressionskoeffizienten von $b = 0,13$. Aufgrund der Additivität des *Linear Mixed Model* (kurz: LMM) bedeutet dies, dass der Einfluss des Vorwissens auf die Qualität von Erklärungen um 0,13 Einheiten verringert wird, wenn in diesen eine optische Repräsentation fehlt. Die Globalnote würde dann im Schnitt nur um 0,12 Einheiten niedriger und damit besser ausfallen. Umgekehrt bedeutet das für den Prädiktor *Erklärmodus.2*, dass dieser nochmals um 0,13 Standardabweichungen im Vorwissen verstärkt wird, wenn die Variable Vorwissen berücksichtigt wird. Das bedeutet, dass die Globalnote um 0,64 Einheiten schlechter bewertet wird, wenn eine optische Repräsentation fehlt und zusätzlich das Vorwissen der Schüler*innen berücksichtigt wird. Dieser Interaktionseffekt kann als Indikator dafür gesehen werden, dass es gerade beim Fehlen einer optischen Repräsentation noch etwas stärker darauf ankommt, was Schüler*innen bereits wissen und über welche musikalischen Erfahrungen sie verfügen. Da die Schüler*innen die fehlende optische Repräsentation umso schlechter bewerten, über je mehr Vorwissen sie verfügen, kann vermutet werden, dass sie aufgrund dessen insbesondere die Darstellung von Notentext gewohnt sind und diese erwarten würden.

Die Varianzaufklärung des Modells beträgt 0,08 % für die festen Effekte und 0,37 % für die festen gemeinsam mit den zufälligen Effekten. Dies kann vor dem Hintergrund der wenigen Prädiktoren als ausreichend bewertet werden, um einen ersten Hinweis auf den Einfluss des Vorwissens der Schüler*innen hinsichtlich der Bewertung von Erklärqualität zu erhalten. Demnach kann *Hypothese 7.1*, dass das Vorwissen der Schüler*innen mit deren Bewertungen zur Erklärqualität zusammenhängt, aufrechterhalten werden.

Erklärmodus als Prädiktor für die Wirksamkeit unterrichtlichen Erklärens im Fach Musik?

Die Kurz-Wissenstests wurden den Schüler*innen nicht nur vor der jeweils ersten Erklärung zu einem Thema, sondern auch im Anschluss daran vorgelegt. Von Interesse war, ob sich bereits in dieser stark verkürzten Form eines *Prä-Post*-Vergleichs erste Hinweise hinsichtlich der Wirksamkeit der drei Erklärmodi zeigen. Allerdings stellt dies keine explizite Forschungsfrage der vorliegenden Arbeit dar – dazu wäre ein umfassenderes Testformat notwendig. Dennoch wurde als *Post-hoc*-Analyse überprüft, ob sich anhand der Erklärungen pro Thema Wissenszuwächse feststellen lassen. Neben dem Prädiktor Zeitpunkt wurde auch der Prädiktor Erklärmodus ins Modell integriert. Es zeigte sich zwar für jedes Thema, dass der Faktor Zeitpunkt signifikant wurde und demnach jede Erklärung zu einem Wissenszuwachs führt. Allerdings konnte kein Einfluss unterschiedlicher Erklärmodi festgestellt werden. Hinsichtlich der Wirksamkeit kann deshalb nur konstatiert werden, dass die Erklärungen scheinbar zu einem Wissenszuwachs bei den Schüler*innen geführt haben. Inwieweit dabei aber Unterschiede im Erklären eine Rolle spielen, kann nicht beantwortet werden.

6 Diskussion

Abschließend werden die Ergebnisse dieser Arbeit in Anbetracht der formulierten Forschungsziele diskutiert. Zunächst erfolgt eine Zusammenfassung der zentralen Ergebnisse, die mit den theoretischen Vorüberlegungen dieser Arbeit in Beziehung gesetzt und in den musikpädagogischen und bildungswissenschaftlichen Forschungsdiskurs eingeordnet werden. Anschließend wird die Aussagenkraft der Untersuchung kritisch reflektiert, indem Einschränkungen aufgrund des forschungsmethodischen Vorgehens benannt sowie weitere Forschungsdesiderata zum Erklären im Musikunterricht aufgezeigt werden. Während auf Basis dessen ein Ausblick auf zukünftige Forschungsfelder erfolgt, schließt dieses Kapitel mit einem Fazit zur vorliegenden Arbeit.

6.1 Zusammenfassung und Einordnung in den Forschungsdiskurs

Ausgangspunkt der vorliegenden Arbeit war die Frage, welche Qualitätsmerkmale unterrichtlichen Erklären im Fach Musik von Bedeutung sind. Sie wurde aufgeworfen, da Erklären einerseits innerhalb der Unterrichtsforschung als Kerntätigkeit einer Lehrkraft aufgefasst wird und fachübergreifend zunehmend in den Fokus des Forschungsinteresses rückt (z. B. Charalambous et al., 2011; Findeisen, 2017; Fraefel & Scheidig, 2018; Leinhardt, 2001; Pauli, 2015) und sich andererseits die Forschungslage zu Qualitätsmerkmalen unterrichtlichen Erklärens im Fach Musik defizitär zeigt (Frei, 2019). Ziel der empirischen Untersuchung war es demnach, Merkmale unterrichtlichen Erklärens im Fach Musik zu identifizieren und deren Bedeutung für gutes Erklären im Fach Musik anhand der Beurteilungen von Schüler*innen, Studierenden und Lehrenden zu untersuchen. Dafür wurde ein Instrument entwickelt, das verschiedene Perspektiven auf Erklären abbilden und sowohl die Bewertung von Erklärqualität als auch Beurteilungen zu ausgewählten Qualitätsmerkmalen erfassen kann.

6.1.1 Erklärqualität

Das erste zentrale Anliegen der Untersuchung bestand darin herauszufinden, ob Unterschiede in den Bewertungen der Erklärqualität auf die Zugehörigkeit zur Statusgruppe und den Einsatz unterschiedlicher Erklärmodi zurückgeführt werden können.

Statusgruppe

Es konnte gezeigt werden, dass die drei Statusgruppen die Qualität von Erklärungen unterschiedlich bewerten. Zwischen Studierenden und Lehrenden unterscheiden sich die absoluten Bewertungen der Erklärungen insofern, dass die Gruppe der Lehrenden sowohl einen größeren Bereich der Notenskala nutzt als auch durchschnittlich schlechtere Noten vergibt. Dadurch sehen sie im Schnitt mehr Potenzial zur Qualitätssteigerung der

Erklärungen als Studierende. Es wird vermutet, dass sich in der stärkeren Differenzierung der Globalurteile ein höheres professionelles Wissen (z. B. Puffer & Hofmann, 2017; siehe für weitere Fächer auch Krauss et al., 2017) beziehungsweise eine höhere Expertise der Lehrenden widerspiegelt (z. B. Gruber & Stöger, 2011; Krauss & Bruckmaier, 2014). Auf deskriptiver Ebene sind über alle Erklärungen hinweg zwischen Studierenden und Lehrenden jedoch ähnliche, in Relation zueinanderstehende Bewertungstendenzen erkennbar, weshalb nicht von einer grundsätzlich divergierenden Vorstellung über die Qualität unterrichtlichen Erklärens im Fach Musik ausgegangen werden sollte.

Die Qualitätsurteile der Schüler*innen unterscheiden sich gegenüber denen der Studierenden und Lehrenden in grundlegenderer Form. Dabei bleiben allerdings Zweifel offen, ob die Bewertungen der Schüler*innen nicht durch andere Aspekte beeinflusst werden, wie es im Kontext der Beurteilung von Unterrichtsqualität festgestellt wurde (z. B. Clausen, 2002; Göllner et al., 2016). So wird angenommen, dass die Bewertung der Erklärqualität auch davon positiv beeinflusst wird, ob Schüler*innen glauben, etwas verstanden zu haben. Dies könnte möglicherweise erklären, warum sie im Gegensatz zu Studierenden und Lehrenden die beiden Erklärungen zum Thema Notenwerte am besten bewertet haben. Denn in der Studie zeigt sich nicht nur, dass Schüler*innen das größte Vorwissen zum Thema Notenwerte besitzen. Sie gaben auch an, die beiden Erklärungen (NW_o; NW_oa) am besten verstanden zu haben. Insgesamt dienen die Unterschiede zwischen den drei Statusgruppen in der Bewertung der Erklärqualität als erster Hinweis darauf, dass eine perspektivenspezifische Validität von Erklärqualität vorliegt, wie sie auch im Kontext von Unterrichtsqualität angenommen wird (siehe Kunter & Baumert, 2006; Kleickmann et al., 2019).

Erklärmodus

Im Gegensatz zum Faktor Statusgruppe hat die Verwendung unterschiedlicher Erklärmodi eine größere Bedeutung für die Bewertung von Erklärqualität. Am besten werden erwartungskonform die Erklärungen bewertet, in denen optische und akustische Repräsentationen miteinander kombiniert werden. Dieses Ergebnis stützt Empfehlungen der Kognitionspsychologie, wonach die Aktivierung mehrerer Kanäle sowohl das Behalten von Inhalten erleichtern als auch Präferenzen unterschiedlicher Lerntypen berücksichtigen kann (z. B. Brown & Hatton, 1982; Findeisen, 2017). Zur Frage, welche Bedeutung die beiden Repräsentationsformen jeweils einzeln betrachtet für die Bewertung der Erklärqualität haben, zeigt sich ein klares Bild: Über alle Statusgruppen hinweg werden vor allem die Erklärungen signifikant schlechter bewertet, in denen eine optische Repräsentation fehlt. Auch wird der Wunsch nach dem Einsatz einer optischen Repräsentation häufiger und intensiver als nach dem Einsatz einer akustischen Repräsentation geäußert. Dieser Befund unterstützt Ergebnisse aus Studien anderer Schulfächer, in denen der Verwendung von Visualisierungen, Grafiken und Abbildungen eine große Bedeutung hinsichtlich der Qualität von Erklärungen beigemessen wird (z. B. Findeisen, 2017; Kulgemeyer, 2016). Auch wenn im musikpädagogischen Diskurs keine expliziten empirischen Befunde darüber vorliegen, welche Bedeutung der Einsatz von Visualisierungen beim

Erklären hat, steht das vorliegende Ergebnis im Einklang mit der Modellierung wichtiger Facetten fachdidaktischen Wissens von Musiklehrkräften (Puffer & Hofmann, 2017).

Zur Bedeutung akustischer Repräsentationen beim Erklären im Musikunterricht lag keine Hypothese vor. Allerdings überraschen die Befunde hinsichtlich der Bedeutung für die Qualität unterrichtlichen Erklärens im Fach Musik. Das Fehlen akustischer Repräsentationen hat über alle Statusgruppen hinweg nur vereinzelte und jeweils schwach ausfallende Effekte auf die Bewertung der Erklärqualität. Zwar werden diese Erklärungen tendenziell etwas schlechter als solche bewertet, die beide Repräsentationsformen einsetzen. Allerdings zeigen sich weder systematische noch signifikante Unterschiede. Während die klangliche Realisation von Musik im Kontext musikunterrichtlicher Kommunikation (Hellberg, 2018) oder als Aspekt fachdidaktischen Wissens von Musiklehrkräften (Puffer & Hofmann, 2017) als zentral und notwendig erachtet wird, zeigt sich eine derartige Bedeutung für die Qualität von Erklärungen nicht. Allerdings beziehen sich die Beurteilungen ausschließlich auf Situationen, in denen musiktheoretische Inhalte erklärt werden und dabei vorrangig kognitive Ziele von Musikunterricht im Zentrum stehen.

So kann insgesamt vermutet werden, dass die Bedeutung des Einsatzes musikbezogener Repräsentationen mit den Zielen zusammenhängt, die im Musikunterricht verfolgt werden. Beim Erklären wird der Einsatz akustischer Repräsentationen zwar auch als wichtig und wünschenswert erachtet, eine zwingende Notwendigkeit, wie sie dem Einsatz optischer Repräsentationen attestiert wird, lässt sich jedoch weder aus den quantitativen noch aus den qualitativen Ergebnissen ableiten.

6.1.2 Qualitätsmerkmale

Das zweite zentrale Anliegen der Studie bestand darin, Qualitätsmerkmale unterrichtlichen Erklärens für das Fach Musik zu identifizieren und ihre Bedeutung hinsichtlich der Erklärqualität zu untersuchen.

Es konnte gezeigt werden, dass alle ausgewählten Merkmale für die Qualität unterrichtlichen Erklärens von Bedeutung sind und in den drei linearen Modellen einen durchaus beachtlichen Teil der Varianz in den Globalurteilen zur Erklärqualität erklären können (siehe Tab. 5.40). Damit liegen erstmals empirisch Erkenntnisse zu Qualitätsmerkmalen unterrichtlichen Erklärens im Fach Musik vor, die sich in Befunde anderer Fachdidaktiken einreihen (z. B. Findeisen, 2017; Schopf & Zwischenbrugger, 2015). Ein weiteres interessantes Ergebnis dieser Arbeit besteht darin, dass jedoch keines der Merkmale in allen Statusgruppen von Bedeutung ist, sondern in jeder Gruppe andere Merkmale als relevant erachtet werden. Daher kann die Vermutung einer perspektivenspezifischen Validität unterrichtlichen Erklärens im Fach Musik weiterhin aufrechterhalten werden.

Zwischen den beiden Perspektiven auf Erklären sind dabei grundsätzliche Unterschiede festzustellen: Während Studierende und Lehrende insbesondere tiefenstrukturelle Merkmale von Erklären wie *Adressatenorientierung* oder *Strukturiertheit* als relevant erachten, sind für Schüler*innen mit dem *Einsatz musikbezogener Repräsentationen* sowie dem *Sprech- und Körperausdruck* der Lehrkraft Merkmale von Bedeutung (siehe Tab. 5.40), die eng mit dem Verhalten der Lehrkraft in Verbindung stehen und demnach eher

auf einer Oberflächenebene anzusiedeln sind. Vor diesem Hintergrund werden für die Beurteilung unterrichtlichen Erklärens einer Lehrkraft ähnliche Muster vermutet, wie sie Wisniewski et al. (2020) für die Bewertung von Unterrichtsqualität zeigen konnten: Schüler*innen fokussieren bei ihren Urteilen vor allem das Verhalten der Lehrkraft und weniger die Inhalte des Unterrichts. Für die Diskussion der einzelnen Merkmale werden nun quantitative wie qualitative Befunde integriert.

Adressatenorientierung

Das Merkmal *Adressatenorientierung* hat hohe Bedeutung für die Qualität einer Erklärung im Fach Musik aus Sicht von Studierenden und Lehrenden. Das zeigt sich sowohl in den Regressionsanalysen als auch in der Inhaltsanalyse der offenen Antworten. Aufgrund der Tatsache, dass *Adressatenorientierung* bei der Gruppe der Lehrenden als einziges, jedoch sehr bedeutsames Merkmal Einfluss auf die Bewertung der Erklärqualität einnimmt, wird vermutet, dass *Adressatenorientierung* für diese Gruppe als Leitprinzip für unterrichtliches Erklären fungiert (z. B. Wittwer & Renkl, 2008) und darin weitere Merkmale subsummiert werden.

Die Auswertungen der qualitativen Daten deuten darauf hin, dass insbesondere die Orientierung am Vorwissen der Schüler*innen für das Erklären im Fach Musik besonders bedeutsam ist. Dieser Befund steht in Einklang damit, dass aufgrund außerschulischer Lernangebote wie Instrumental- oder Gesangsunterricht zwischen den Schüler*innen im Fach Musik eine große Heterogenität ihrer Kenntnisse sowie ihrer musikalischen Vorerfahrungen vorliegt (z. B. Jank, 2013; Linn, 2017; Niessen, 2013). Gleichzeitig reiht er sich auch in Ergebnisse anderer fachdidaktischer Studien zum unterrichtlichen Erklären ein, in denen die Orientierung am Vorwissen der Schüler*innen als relevant erachtet wird (z. B. Findeisen, 2017; Kulgemeyer & Tomczyszyn, 2015; Schopf & Zwischenbrugger, 2015).

Zudem können die Ergebnisse der inhaltsanalytischen Auswertung als Hinweis dafür dienen, dass die Orientierung an der Lebenswelt der Schüler*innen die Qualität von Erklärungen im Fach Musik erhöht, wie es sich beispielsweise auch im Fach Wirtschaft zeigt (z. B. Findeisen, 2017; Schopf & Zwischenbrugger, 2015). So können im Fach Musik die Vorlieben und Erfahrungen der Schüler*innen berücksichtigt werden (z. B. Kaiser, 2010; Richter, 1993). Die Bedeutung didaktischer Reduktion, wie sie beispielsweise von Oebelsberger (2019) für die Vermittlung musiktheoretischer Inhalte als relevant erachtet wird, kann durch die vorliegenden Daten nicht in gleicher Form bestätigt werden.

Strukturiertheit

Die *Strukturiertheit* einer Erklärung hat insbesondere bei den Studierenden eine hohe Bedeutung für die Qualität unterrichtlichen Erklärens im Fach Musik. Während das Merkmal in den Regressionsmodellen der Schüler*innen und Lehrkräfte keinen signifikanten Einfluss zeigt, deuten in beiden Gruppen dennoch verschiedene Aspekte auf einen

wichtigen Stellenwert hin: Bei den Schüler*innen korrelieren die Beurteilungen zum Merkmal *Strukturiertheit* hoch mit denen zum *Einsatz musikbezogener Repräsentationen*; dieses Merkmal wiederum zeigt im Regressionsmodell einen signifikanten Einfluss auf die Erklärqualität (siehe Tab. 5.34). Es wird daher vermutet, dass der *Einsatz musikbezogener Repräsentationen* auch strukturierende Funktion einnehmen kann.

Die Lehrenden (sowie auch die Studierenden) hingegen begründen mit Aussagen zur *Strukturiertheit* ihre Bewertungen der Erklärqualität, was auch auf gewisse Relevanz des Merkmals hindeutet. Dabei treten zwei Aspekte hervor: Zum einen unterstützt insbesondere der Befund, dass sach- beziehungsweise chronologisch aufgebaute Erklärungen im Musikunterricht zur Qualitätssteigerung beitragen, Ergebnisse aus der Musikdidaktik (z. B. Jank & Meyer, 2013) und anderen Fachdidaktiken (z. B. Schopf & Zwischenbrugger, 2015). Zum anderen wird die Frage verhandelt, an welcher Stelle beim Erklären musiktheoretischer Inhalte Musik erklingen sollte. Dabei kristallisiert sich die Tendenz heraus, zuerst Prinzipien und Konzepte zu erklären, bevor eine klangliche Demonstration erfolgt. Damit kann dieser Befund einerseits die Empfehlung von Fuchs (2015), theoretische Erklärungen erst dann zu geben, wenn Schüler*innen über ausreichend musikalische Erfahrungen verfügen, pauschal nicht unterstützen. Andererseits steht er in Einklang mit Empfehlungen von Renkl et al. (2006) aus dem Bereich der Informatik, den Fokus insbesondere dann auf prinzipienorientierte Erklärungen zu richten, wenn Regeln oder Schemata erkannt und abstrahiert werden sollen, wie es auch für das Verstehen musiktheoretischer Inhalte notwendig sein kann. Hinweise in den offenen Antworten, dass solche Erklärungen zu Beginn auch daran anschließende Phasen konstruktivistischen Lernens einleiten können, decken sich ebenfalls mit Empfehlungen aus anderen Fachdidaktiken (siehe Leisen, 2007).

Sprachliche Verständlichkeit

Die Relevanz des Merkmals *Sprachliche Verständlichkeit*, wie sie für das Erklären in anderen Fachdidaktiken vorliegt (z. B. Schopf & Zwischenbrugger, 2015), kann durch die Befunde dieser Arbeit für das Fach Musik nur bedingt unterstützt werden. Zum einen unterliegt die Interpretation der inferenzstatistischen Ergebnisse einigen Limitationen: Während die Skala der Schüler*innen eine niedrige interne Konsistenz aufweist, erscheinen die Befunde der Studierenden aufgrund negativer Effekte kontraintuitiv. Zum anderen hat das Merkmal aus Sicht der Lehrenden keinerlei Bedeutung für die Erklärqualität im Fach Musik (siehe Tab. 5.39). Allerdings wird angenommen, dass die Ergebnisse weniger durch Spezifika des Fachs Musik zu begründen, sondern vielmehr auf die Form der Operationalisierung zurückzuführen sind. Diese weist aufgrund der breiten Erfassung einen formativen Charakter auf, der für eine Kombination mit sprachwissenschaftlichen Analysen zwar sinnvoll erscheint (siehe Thim-Mabrey & Lindl, i. Vorb.), sich jedoch in dieser Form nur eingeschränkt als Prädiktor für das Kriterium Erklärqualität eignet.

Aus der Analyse der offenen Antworten geht hervor, dass die angemessene Verwendung von Fachbegriffen größere Bedeutung zu haben scheint, was im Einklang mit Befunden anderer fachdidaktischer Studien zu Qualitätsmerkmalen unterrichtlichen

Erklärens steht (z. B. Drollinger-Vetter, 2011; Findeisen, 2017; Kulgemeyer & Tomczyszyn, 2015). Allerdings divergieren die Einschätzungen zur angemessenen Verwendung, sodass anhand der vorliegenden Daten keine Empfehlungen zum Umgang mit Fachbegriffen beim Erklären im Fach Musik herausgearbeitet werden können.

Sprech- und Körperausdruck

Die Bedeutung des Merkmals *Sprech- und Körperausdruck* zeigt sich in dieser Studie nur für die Gruppe der Schüler*innen. Wie bereits dargelegt, muss dieser Befund jedoch vor dem Hintergrund interpretiert werden, dass sich die Beurteilungen der Schüler*innen in der Tendenz vor allem auf Oberflächenmerkmale unterrichtlichen Erklärens fokussieren. Bei Studierenden und Lehrenden zeigt sich weder in den Regressions- noch in den Inhaltsanalysen, dass *Sprech- und Körperausdruck* eine Bedeutung für die Erklärqualität hat. Damit können die Befunde fachdidaktischer Studien aus anderen Disziplinen nur für die Perspektive der Schüler*innen unterstützt werden (z. B. Schopf & Zwischenbrugger, 2015).

Beim Blick auf die deskriptiven Werte der Beurteilung des Merkmals *Sprech- und Körperausdruck* fällt auf, dass sowohl über alle Studienteilnehmer*innen als auch getrennt nach Statusgruppen zwischen einzelnen Erklärungen große Unterschiede vorliegen (siehe Tab. 5.29 bis 5.33). Dies überrascht insbesondere deshalb, weil in allen Videos dieselbe Lehrkraft erklärt. Es besteht daher die Annahme, dass Einschätzungen zum Sprech- und Körperausdruck nicht nur auf die Performanz der Lehrkraft zurückgeführt werden können. Möglicherweise sind diese auch vom jeweiligen Erklärthema oder dem eingesetzten Erklärmodus abhängig, beziehungsweise hängen mit sprachlichen Aspekten der Erklärung zusammen. Die Vermutung, dass nicht nur personale Merkmale relevant sind, kann auch dadurch gestützt werden, dass *Sprech- und Körperausdruck* über alle Statusgruppen hinweg erwartungsgemäß hoch mit der Kontrollvariable *Persönlichkeitswirkung* korreliert ist, diese jedoch keine signifikante Bedeutung für die Erklärqualität besitzt (siehe Tab. 5.40).

Einsatz musikbezogener Repräsentationen

Dass der *Einsatz musikbezogener Repräsentationen* für die Qualität unterrichtlichen Erklärens im Fach Musik von hoher Bedeutung ist, kann auf unterschiedlichen Ebenen gezeigt werden. Ergänzend zu den Befunden, die aus der systematischen Variation optischer und akustischer Repräsentationsformen hervorgehen, zeigt die Beurteilung der Skala im Regressionsmodell der Schüler*innen einen signifikanten Einfluss auf die Bewertung der Erklärqualität. Während dieser bei Studierenden und Lehrenden nicht vorliegt, kann jedoch die hohe Zahl an Aussagen innerhalb der offenen Antworten darauf hindeuten, dass das Merkmal für beide Gruppen dennoch hinsichtlich der Erklärqualität von Bedeutung ist.

Bezüglich der verschiedenen Formen von Repräsentation tragen die offenen Antworten zur Differenzierung der Befunde bei. Hinsichtlich des Einsatzes optischer Repräsentationen wird zum einen die Abbildung von Notentext als Notwendigkeit für das Erklären im Fach Musik deklariert, wie es bei der Vermittlung musiktheoretischer Inhalte bereits lange Tradition ist (Ehrenforth, 2010; Gruhn, 2003a). Darüber hinaus wird auch die Bedeutung von Visualisierungen herausgehoben, die als Hilfsmittel zur Veranschaulichung eingesetzt werde – Jünger (2006) konnte dies bereits im Zuge der Verwendung von Schulbüchern für den Musikunterricht beobachten. Dabei divergieren die Einschätzungen sowohl hinsichtlich des Einsatzes der Klaviatur (siehe dazu Beiderwieden, 2008) als auch bezüglich der pyramidalen Darstellung des Verhältnisses einzelner Notenwerte zueinander, sodass aus beiden Beispielen keine klaren Empfehlungen abzuleiten sind. Große Übereinstimmung zeigt sich jedoch bezüglich der Notwendigkeit optischer Repräsentationen insbesondere für das Verstehen von Schüler*innen mit wenig Vorwissen und fehlender musikalischer Vorerfahrung.

Wie bereits dargelegt, können diese kognitiven Ziele möglicherweise erklären, warum dem Einsatz akustischer Repräsentationen in Form von Hör- oder Klangbeispielen weniger Bedeutung als der Verwendung optischer Repräsentationen beigemessen wird. Akustische Repräsentationen werden als wichtig erachtet, um das Phänomen Musik klanglich zu veranschaulichen (z. B. Hellberg, 2018). Allerdings werden sie insbesondere für Schüler*innen mit fehlendem Vorwissen und mangelnder musikalischer Vorerfahrung dann als nicht hinreichend deklariert, wenn sie ohne optischen Darstellung beim Erklären zum Einsatz kommen. Aus diesen Beobachtungen ergibt sich als Forschungsfrage für zukünftige Vorhaben, wie wirksam der Einsatz akustischer Repräsentationen für den Lernzuwachs von Schüler*innen insbesondere vor dem Hintergrund heterogener Lernvoraussetzungen ist. Antworten darauf könnten den bereits seit längerer Zeit geführten musikpädagogischen Diskurs zum Umgang mit Heterogenität (z. B. Linn, 2017; Niessen, 2013; Puffer, 2022) um einen wichtigen Aspekt ergänzen.

Im Gegensatz zu optischen und akustischen spielen weitere Formen der Repräsentation in den offenen Antworten eine nachgeordnete Rolle. Hervorzuheben ist als Maßgabe für unterrichtliches Erklären im Fach Musik, beim Einsatz von Analogien insbesondere die Aspekte klar zu benennen, die mit dem zu erklärenden Zielkonzept nicht übereinstimmen. Dies konnten bereits Harrison und Treagust (1993) für den Einsatz von Analogien im Physikunterricht zeigen. Die Hinweise zum Einsatz von (Musik-)Beispielen beziehen sich insbesondere auf die Technik der Verfremdung. Dabei können keine Aussagen abstrahiert werden, die den Einsatz von Beispielen beim Erklären im Musikunterricht näher spezifizieren.

Weitere potenzielle Qualitätsmerkmale

Anhand der offenen Antworten kann insbesondere die fachliche Korrektheit einer Erklärung im Fach Musik als weiteres Qualitätsmerkmal vorgeschlagen werden. Dieses Erkenntnis steht im Einklang mit Befunden anderer Fächer wie Wirtschaft oder Mathematik (z. B. Findeisen, 2017; Wagner & Wörn, 2011). Schilcher et al. (2017) bewerten die

Notwendigkeit fachlicher Korrektheit beim Erklären als trivial. Dies kann auch für das Fach Musik vermutet werden. Darüber hinaus können auf Basis der vorliegenden qualitativen Daten keine weiteren Merkmale unterrichtlichen Erklärens im Fach von umfassender Bedeutung festgestellt werden.

6.2 Einschränkungen und Ausblick

6.2.1 Forschungsmethodisches Vorgehen

Durch das forschungsmethodische Vorgehen war es möglich, verschiedene Perspektiven auf Qualitätsmerkmale unterrichtlichen Erklärens abzubilden. Auch konnten sowohl explorative als auch hypothesenprüfende Fragestellungen umgesetzt werden. Vorteilhaft erwies sich insbesondere, dass alle Statusgruppen *fremden Unterricht* bewertet haben, da die Ergebnisse aufgrund derselben Ausgangspositionen und Voraussetzungen gut vergleichbar waren. Auch konnte dadurch die subjektive Involviertheit von Schüler*innen und Lehrkräften, die in anderen Studien eigenen Unterricht beurteilen, ausgeschlossen werden (z. B. Kleickmann et al., 2019). Dies äußert sich anhand des statistisch unbedeutenden Einflusses der Kontrollvariable *Persönlichkeitswirkung* in allen Statusgruppen. Das Verfahren, aus verschiedenen Perspektiven *fremden Unterricht* zu beurteilen, kann damit auch für zukünftige Forschungsvorhaben zur Beschreibung von Erklär- oder Unterrichtsqualität empfohlen werden.

Allerdings gehen aus dem methodischen Vorgehen auch Einschränkungen hervor. Aufgrund der eher kleineren Stichprobengröße ($N_{\max} = 271$) ist die externe Validität der Ergebnisse grundsätzlich begrenzt. Zudem wurde ein Großteil der Studierenden und Lehrenden über berufliche oder persönliche Kontakte angesprochen, weshalb keine repräsentative Stichprobe vorliegt. Indem die Statusgruppe der Lehrenden aus Lehrkräften und Didaktiker*innen besteht, die zum Zeitpunkt der Untersuchung in unterschiedlichen Bereichen (Schule und Universität beziehungsweise Hochschule) tätig waren, sind die Ergebnisse für diese Statusgruppe in ihrer Differenziertheit beeinträchtigt. Eine Unterscheidung beider Gruppen, wie sie an anderer Stelle unternommen wurde – jedoch mit erheblichen Limitationen aufgrund der Stichprobengröße (siehe Frei et al., i. Vorb.) – wird für nachfolgende Untersuchungen empfohlen.

Das Design der Studie erfordert in der Durchführung einen erheblichen zeitlichen Aufwand. So wird einerseits angenommen, dass dies unter anderem zu einer eingeschränkten Teilnahmebereitschaft bei Studierenden und Lehrenden geführt hat. Andererseits wird vermutet, dass die 40-minütige Erhebungsdauer für die beteiligten Schüler*innen zu Konzentrationsmängeln am Ende des Fragebogens geführt haben könnte. Bei dieser Statusgruppe wurden aufgrund dieser ohnehin hohen zeitlichen Anforderung keine Items mit offenen Antwortformaten gestellt, weshalb diese wertvollen Daten für die Inhaltsanalyse fehlen. Um diesem Desiderat in zukünftigen Forschungsvorhaben zu begegnen, werden kognitive Interviews unter Anwendung der Probing-Technik als Form der Erhebung vorgeschlagen. Sie haben sich innerhalb der Pilotierung als fruchtbare Methode

erwiesen, um differenzierte individuelle Sichtweisen von Schüler*innen zu erfassen (siehe Frei, 2019). Diese ausführliche Erhebung war im Rahmen der vorliegenden Studie nicht möglich, wird jedoch für weitere Forschungsvorhaben nachdrücklich empfohlen.

Die Analyse der offenen Antworten unterliegt verschiedenen Einschränkungen. Zunächst impliziert die Auswertung qualitativer Daten, dass die Ergebnisse nicht verallgemeinerbar sind. Als problematisch hat sich die Bitte erwiesen, dass die Studienteilnehmer*innen *maximal drei kurze Stichpunkte* als Begründung zum Globalurteil abgeben sollten. Da einige Antworten nur einzelne oder wenige Wörter umfassten, war die Interpretation der Aussagen erschwert und die Auswertungsobjektivität zusätzlich beeinträchtigt. Für weitere Forschungsvorhaben wäre die Empfehlung, um vollständige Sätze bei der Beantwortung der Items zu bitten. Auch sollten die offenen Antworten dahingehend unterschieden werden, dass explizit sowohl nach positiven und negativen Gründen für die Bewertung der Erklärqualität gefragt wird. Es wird vermutet, dass dadurch die Aussagekraft zur Erklärqualität gesteigert werden könnte. Allerdings ist auch zu befürchten, dass die Studienteilnehmer*innen dadurch verstärkt nach positiven wie negativen Begründungen suchen und so ihre Begründung zur Bewertung der Erklärqualität äußeren Einflüssen unterliegt.

Hinsichtlich der durchgeführten statistischen Analysen sind insbesondere die Ergebnisse der *mixed ANOVA* aufgrund teils fehlender Varianzhomogenität zwischen den Gruppen mit Einschränkungen in der Aussagekraft behaftet. Es empfiehlt sich für weitere Analysen, auf regressionsanalytische Verfahren zu setzen, da diese weniger strenge Modellvoraussetzungen verlangen (z. B. Hilbert et al., 2019). Auch erfordern die niedrigen internen Konsistenzen der Skala *Sprachliche Verständlichkeit*, deren bisherige Operationalisierung zu überprüfen und das Merkmal insgesamt enger zu erfassen, sodass ein reflexives Konstrukt entsteht. Gestaltet sich dies aus inhaltlichen Gründen schwierig, wäre auch zu überlegen, ob die einzelnen Facetten des Merkmals grundsätzlich anders konzeptualisiert werden können: Während die Angemessenheit der Verwendung von Fachbegriffen sich immer auch an den jeweiligen Adressat*innen orientiert und demnach auch dem Merkmal *Adressatenorientierung* zugeordnet werden könnte, kann die sprachliche Akzentuierung wichtiger Hinweise eine strukturierende Funktion einnehmen, weshalb dieses Item auch zum Merkmal *Strukturiertheit* passen könnte.

Nachdem für unterrichtliches Erklären im Fach Musik zum ersten Mal Qualitätsmerkmale konzeptualisiert und operationalisiert wurden, sind Weiterentwicklungen selbstredend für alle hier vorliegenden Merkmale erforderlich und im Zuge weiterer Forschungsvorhaben wünschenswert. Gerade mit Blick auf die Erhebung von Schüler*innen einer sechsten Jahrgangsstufe könnten auch die Formulierungen der Items nochmals auf Verständnisschwierigkeiten hin überprüft werden. Denn es war auffällig, dass die drei negativ korrelierten Items (St_4, SV_2 und Ad_1; siehe Kap. 5.2.1) jeweils das Intensitätspartikel *zu* verwenden, weshalb die gegenläufigen Einschätzungen der Schüler*innen unter anderem auch darauf zurückgeführt werden könnten.

6.2.2 Erklären im Musikunterricht

Die vorliegende Arbeit konnte zeigen, dass Qualitätsmerkmale unterrichtlichen Erklärens im Fach Musik vorliegen, die zum einen Parallelen zum fachübergreifend geführten Diskurs um unterrichtliches Erklären aufweisen und zum anderen auch der fachspezifischen Besonderheit von Musik als klingendem Phänomen Rechnung tragen. Auch wurde deutlich, dass die ausgewählten Qualitätsmerkmale einer perspektivenspezifischen Validität unterliegen, weshalb die verschiedenen Perspektiven auf unterrichtliches Erklären auch in zukünftigen Forschungsvorhaben berücksichtigt werden sollten.

Um das Konstrukt unterrichtliches Erklären vor diesem Hintergrund noch sorgfältiger und differenzierter zu untersuchen, könnte die strukturelle Validität zum einen in Form einer konfirmatorischen Faktorenanalyse im Mehrgruppenansatz überprüft werden. Diese Analyse kann zeigen, inwieweit das Konstrukt Erklären aus den unterschiedlichen Perspektiven beziehungsweise Statusgruppen anhand derselben Faktoren (Merkmale guten Erklärens) abgebildet werden kann. Zum anderen könnte dabei mit Blick auf das Modell guten Erklärens beziehungsweise guten Instruierens im Musikunterricht (Frei et al., 2022) untersucht werden, ob das Merkmal *Adressatenorientierung* auch statistisch als Leitprinzip unterrichtlichen Erklärens und demnach auf einer übergeordneten Ebene abgebildet werden kann. Neben diesen grundsätzlichen Empfehlungen ergeben sich weitere Anregungen für Forschung zu Qualitätsmerkmalen unterrichtlichen Erklärens im Fach Musik aus Limitationen der vorliegenden Arbeit.

So ist die ökologische Validität der Studie aus unterschiedlichen Gründen begrenzt. Zwar wurden unterrichtsnahe Situationen nachgestellt, allerdings können daraus nur unter Vorbehalt Rückschlüsse auf Erklären im *normalen* Musikunterricht gezogen werden. Weitere Forschungsvorhaben sollten daher Merkmale guten Erklärens in realen Unterrichtssituationen untersuchen. Auch entstehen Einschränkungen aufgrund des experimentellen Designs: Um auf das Erklären der Lehrkraft zu fokussieren, wurden nur monologische Erklärungen ohne Zwischenfragen der Schüler*innen untersucht – eine solche Laborsituation entspricht nicht der Unterrichtsrealität. So wäre zu überlegen, in weiteren Forschungsvorhaben auch den Aspekt spontaner, adaptiver Erklärungen aufzugreifen, mit denen eine Lehrkraft eher unvorbereitet beispielsweise auf Nachfragen von Schüler*innen reagieren muss. Auch wurde das Handeln der Lehrkraft durch die Vorgaben zum Einsatz der Erklärmodi künstlich begrenzt. Das wird beispielsweise dadurch deutlich, dass die Lehrkraft bei der Produktion von Erklärung TL_a erwähnte, dieses Thema im realen Musikunterricht nicht in der vorgegebenen Form, also ohne den Einsatz optischer Repräsentationen, zu erklären.

Mit der Vermittlung musiktheoretischer Inhalte wurde ein bestimmter Bereich von Musikunterricht ausgewählt. Dieser zeigte sich passend, um anhand dessen Qualitätsmerkmale unterrichtlichen Erklärens im Fach Musik modellieren zu können. Im Anschluss an diese Arbeit wäre es jedoch erforderlich, die Passung der Modellierung auch für andere Bereiche des Musikunterrichts zu überprüfen, beispielsweise wenn im Zuge von Singesituationen erklärt wird, welche stimmphysiologischen Vorgänge sich dabei vollziehen. Zudem könnte musikpraktisches Handeln der Schüler*innen im Kontext

musikunterrichtlichen Erklärens in zukünftigen Forschungsvorhaben berücksichtigt werden, da dieses in der vorliegenden Studie aus forschungsmethodischen wie -ökonomischen Gründen ausgespart wurde. Zusätzlich könnten auch taktile Formen der Repräsentation als Gegenstand der Untersuchung dienen. Sie wurden zwar in das Modell guten Erklärens und guten Instruierens im Musikunterricht integriert, eine empirische Überprüfung hinsichtlich ihrer Bedeutung für gutes Erklären beziehungsweise gutes Instruieren steht bisher allerdings noch aus (siehe Frei et al., 2022; siehe Kap. 2.4.2).

Vor dem Hintergrund solcher Unterrichtsinhalte, in denen Musikpraxis eine zentrale Rolle einnimmt, wäre außerdem ein wichtiges Anliegen für zukünftige Forschungsbemühungen, das Verhältnis zwischen Erklären und Instruieren im Musikunterricht in den Fokus zu nehmen. Dabei sollte zunächst die Vermutung empirisch überprüft werden, inwieweit die Modellierung der Qualitätsmerkmale guten Erklärens auch für gutes Instruieren Gültigkeit erlangt (siehe Frei et al., 2022). Auf Basis dessen könnten dann auch Wechselwirkungen zwischen diesen nahe verwandten, jedoch auf unterschiedliche Zwecke abzielenden Unterrichtshandlungen von Musiklehrkräften untersucht werden.

Abschließend wären Forschungsergebnisse zur Wirksamkeit unterrichtlichen Erklärens im Fach Musik von Interesse. Dabei sollte zum einen untersucht werden, ob sich die Erklärmodi, die in der vorliegenden Untersuchung mit einer besseren Bewertung der Erklärqualität einhergehen, auch hinsichtlich eines Kompetenzzuwachses der Schüler*innen als wirksamer erweisen. Anhand der Ergebnisse aus der vorliegenden Studie wäre zu erwarten, dass das Fehlen einer optischen Repräsentationen mit signifikant geringeren Kompetenzzuwächsen der Schüler*innen einhergeht. Während die vorliegende Arbeit vor diesem Hintergrund auch erste Hinweise über den Einfluss des Vorwissens der Schüler*innen auf die Bewertung von Erklärqualität liefern kann – insbesondere dann, wenn optische Repräsentationsformen fehlen –, sind aufgrund der sparsamen Testverfahren keine Aussagen über die Wirksamkeit der Erklärmodi möglich. Zum anderen könnte ein Ziel zukünftiger Forschungsvorhaben darin liegen, die Wirksamkeit der einzelnen Qualitätsmerkmale unterrichtlichen Erklärens zu untersuchen, um so Aussagen über deren jeweilige prädiktive Validität treffen zu können.

6.3 Fazit

Der wissenschaftliche Ertrag dieser empirischen Untersuchung liegt unter anderem darin, dass für das Fach Musik erstmals Qualitätsmerkmale unterrichtlichen Erklärens identifiziert und erste Erkenntnisse darüber gewonnen werden konnte, welche Bedeutung sie für die Qualität von Erklären im Musikunterricht haben. Die Befunde sollen auch dazu dienen, den Diskurs über gutes Erklären im Fach Musik in Gang zu setzen.

Es konnte gezeigt werden, dass *Adressatenorientierung*, *Strukturiertheit* und *Einsatz musikbezogener Repräsentationen* sowie mit Einschränkungen auch *Sprachliche Verständlichkeit* und *Sprech- und Körperausdruck* der Lehrkraft relevante Qualitätsmerkmale unterrichtlichen Erklärens im Fach Musik darstellen. Auch konnten erste Hinweise auf Wechselbeziehungen zwischen einzelnen Merkmalen herausgearbeitet werden. Die

Ergebnisse dieser Arbeit legen zudem nahe, eine perspektivenspezifische Validität von Erklärqualität im Fach Musik anzunehmen: Zum einen bewerten die drei Statusgruppen die Qualität unterrichtlichen Erklärens unterschiedlich. Zum anderen sind für alle drei Gruppen jeweils verschiedene Merkmale hinsichtlich der Qualität unterrichtlichen Erklärens im Fach Musik von Bedeutung.

Darüber hinaus wurde deutlich, dass insbesondere das Fehlen einer optischen Repräsentation bei der Vermittlung musiktheoretischer Inhalte zu deutlich schlechteren Bewertungen der Erklärqualität führt. Es wird daher vermutet, dass optische Repräsentationen beim Erklären im Musikunterricht gerade für das Erreichen kognitiver Unterrichtsziele von größerer Bedeutung als akustische Repräsentationen sind. Erste Anzeichen deuten darauf hin, dass das Vorwissen der Schüler*innen zu Unterschieden hinsichtlich der Bewertung von Erklärqualität im Fach Musik führt.

Auch wurde im Zuge dieser Arbeit ein Onlinefragebogen als Erhebungsinstrument entwickelt, der zukünftigen Forschungsvorhaben eine Grundlage darstellen kann. Darin dienen kurze Videovignetten als Stimuli, um Bewertungen der Erklärqualität und Beurteilungen über ausgewählte Qualitätsmerkmale zu erfassen. Aufgrund verschiedener Varianten ist der Fragebogen in der Lage, unterschiedliche Perspektiven auf Erklären zu berücksichtigen und Vergleiche zwischen diesen zu ermöglichen. Dabei kann anhand der Erhebung quantitativer wie qualitativer Daten das Konstrukt Erklären ausführlich erfasst werden. Zudem misst ein Wissenstest in knapper Form das Vorwissen der Schüler*innen zu den Themen Notenwerte, Taktarten und Tonleiter.

In den theoretischen Ausführungen der vorliegenden Arbeit wurden Verbindungslinien zwischen Qualitätsmerkmalen (musik-)unterrichtlichen Erklärens und Merkmalen guten (Musik-)Unterrichts herausgearbeitet sowie Bezüge zum Diskurs um professionelle Kompetenzen von Musiklehrkräften hergestellt. Damit besteht die Hoffnung, dass die Ergebnisse dieser Arbeit auch zu diesen beiden musikpädagogischen Diskursen einen Beitrag leisten können. Gleichzeitig ist damit auch der Wunsch verbunden, die Diskussionen um Unterrichts- und Erklärqualität im Fach Musik beziehungsweise zu professionellen Kompetenzen von Musiklehrkräften zu intensivieren und dabei gleichermaßen fachliche Spezifika wie fachübergreifende Ansätze in den Blick zu nehmen. Denn gerade die transdisziplinäre Herangehensweise dieser Studie hat sich als äußerst gewinnbringend erwiesen.

Weiterhin ermöglichen die Befunde der vorliegenden Arbeit auch Implikationen für die musikunterrichtliche Praxis sowie für die Aus- und Weiterbildung von Musiklehrkräften. Die erstellten Videovignetten dienen bereits in positiv bewerteten Seminarkonzepten zum guten Erklären innerhalb des Lehramtsstudiums im Fach Musik als Reflexionsanlass (Ehras et al., 2021). Daher könnten solche Formate beispielsweise auch in der zweiten oder dritten Phase der Lehrkräfteausbildung zum Einsatz kommen. Darüber hinaus können die identifizierten Qualitätsmerkmale auch als Analyseraster für unterrichtliches Erklären und Instruieren im Fach Musik dienen und zur Planung, Durchführung und Reflexion von Erklärungen im Musikunterricht herangezogen werden (Frei et al., 2022).

Diese Arbeit wurde mit einem Zitat von Pauli (2015) eingeleitet, in dem Erklären fachübergreifend als Kerntätigkeit einer Lehrkraft und gute Erklärungen als

unverzichtbares Element von Unterricht deklariert wurden. Anhand der Ergebnisse der vorliegenden Arbeit lautet das Plädoyer nicht nur, Erklären als Kerntätigkeit einer Musiklehrkraft anzusehen, sondern auch, dass gute Erklärungen im Musikunterricht unverzichtbar sind – zumindest, wenn darin kognitive Ziele verfolgt werden.

Abbildungsverzeichnis

Abbildung 2.1	Erklären ₁ und Erklären ₂	20
Abbildung 2.2	Modell guten Erklärens bzw. Instruierens im Musikunterricht	95
Abbildung 3.1	Zusammensetzung der Statusgruppen	115
Abbildung 4.1	Forschungsdesign	125
Abbildung 4.2	Wissenstest Notenwerte	135
Abbildung 5.1	Profilplots zu den arithmetischen Mittelwerten der Merkmale unterrichtlichen Erklärens getrennt nach Statusgruppen	161
Abbildung 5.2	Profilplots zu den Globalurteilen der Faktoren Erklärmodus und Statusgruppe (Thema Notenwerte)	182
Abbildung 5.3	Profilplots zu den Globalurteilen der Faktoren Erklärmodus und Statusgruppe (Thema Taktarten)	183
Abbildung 5.4	Profilplots zu den Globalurteilen der Faktoren Erklärmodus und Statusgruppe (Thema Tonleiter)	184
Abbildung 5.5	Durchschnittliche Globalurteile der Schüler*innen getrennt nach den drei Versionen	186
Abbildung 5.6	Durchschnittliche Globalurteile der Studierenden getrennt nach den drei Versionen	188
Abbildung 5.7	Durchschnittliche Globalurteile der Lehrenden getrennt nach den drei Versionen	189
Abbildung 5.8	Globalurteile pro Video getrennt nach Statusgruppen	191

Tabellenverzeichnis

Tabelle 4.1	Überblick der Gesamtstichprobe	127
Tabelle 4.2	Übersicht zu den Inhalten der Videovignetten	129
Tabelle 4.3	Reihenfolge der Videovignetten	129
Tabelle 4.4	Operationalisierung Sprech- und Körperausdruck	136
Tabelle 4.5	Operationalisierung Einsatz musikbezogener Repräsentationen	138
Tabelle 4.6	Operationalisierung Sprachliche Verständlichkeit	140
Tabelle 4.7	Operationalisierung Strukturiertheit	141
Tabelle 4.8	Operationalisierung Adressatenorientierung	141
Tabelle 4.9	Operationalisierung Persönlichkeitswirkung	142
Tabelle 4.10	Operationalisierung Zusätzliche Kontrollvariablen	143
Tabelle 4.11	Offene Itemformate	144
Tabelle 4.12	Anzahl der offenen Antworten über alle Videos	157
Tabelle 4.13	Anzahl der Zeichen in den offenen Antworten über alle Videos	158
Tabelle 5.1	Bayesianische <i>t</i> -Tests pro Video zwischen den Statusgruppen Lehrkräfte und Didaktiker*innen	160
Tabelle 5.2	Deskriptive Werte der Beurteilungen zu Merkmalen guten Erklärens sowie der Kontrollvariable Persönlichkeitswirkung	160
Tabelle 5.3	Interne Konsistenzen der erhobenen Skalen für Schüler*innen (<i>N</i> = 148) (Cronbachs <i>α</i> und <i>part-whole</i> -korrigierte mittlere Itemtrennschärfen r_{it} sowie deren Range)	163
Tabelle 5.4	Interne Konsistenzen der erhobenen Skalen für Studierende (<i>N</i> = 61) (Cronbachs <i>α</i> und <i>part-whole</i> -korrigierte mittlere Itemtrennschärfen r_{it} sowie deren Range)	165
Tabelle 5.5	Interne Konsistenzen der erhobenen Skalen für Lehrende (<i>N</i> = 62) (Cronbachs <i>α</i> und <i>part-whole</i> -korrigierte mittlere Itemtrennschärfen r_{it} sowie deren Range)	166
Tabelle 5.6	Arithmetisches Mittel und Standardabweichung von Item „Kv_4“	168
Tabelle 5.7	Interne Konsistenzen der drei Wissenstests zu den Themen Notenwerte, Taktarten und Tonleiter für die Gruppe der Schüler*innen (<i>N</i> = 142) für den ersten und zweiten Messzeitpunkt (Cronbachs <i>α</i> und <i>part-whole</i> -korrigierte mittlere Itemtrennschärfen r_{it} sowie deren Range)	170
Tabelle 5.8	Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu allen Codes	172
Tabelle 5.9	Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu den Codes des Bereichs <i>Überzeugungen</i>	173
Tabelle 5.10	Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu den Codes des Bereichs <i>Qualitätsmerkmale I</i>	174

Tabelle 5.11	Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu den Codes des Bereichs <i>Qualitätsmerkmale II</i>	174
Tabelle 5.12	Werte der prozentualen Übereinstimmung (70 %) sowie Cohens Kappa zu den Codes des Bereichs <i>themenbezogene Kategorien</i>	175
Tabelle 5.13	Mittelwerte und Standardabweichungen (getrennt nach Erklärmodus, Thema und Statusgruppe) der Globalurteile zur Erklärqualität	178
Tabelle 5.14	Mittelwerte (Standardabweichungen) der Globalurteile pro Video und Statusgruppe; je Erklärthema eine Varianzanalyse mit gemischtem Design sowie den Faktoren Erklärmodus und Statusgruppe	181
Tabelle 5.15	Kontrollvariablen mit Einschätzungen der Schüler*innen zu ihrem themenbezogenen Vorwissen und videobezogenen Verstehen	187
Tabelle 5.16	Einschätzungen zum Verstehenspotenzial der Erklärung für Schüler*innen durch Studierende	188
Tabelle 5.17	Einschätzungen zum Verstehenspotenzial der Erklärung für Schüler*innen durch Lehrende	190
Tabelle 5.18	Mittelwerte und Standardabweichungen zu allen Erklärungen Getrennt nach Statusgruppen sowie Ergebnisse der <i>Bonferroni</i> -korrigierten <i>Post-hoc</i> -Analysen	192
Tabelle 5.19	Effektstärken zu den Mittelwertsunterschieden zwischen Studierenden und Lehrenden	192
Tabelle 5.20	Effektstärken zu den Mittelwertsunterschieden zwischen Schüler*innen und Lehrenden	193
Tabelle 5.21	Effektstärken zu den Mittelwertsunterschieden zwischen Schüler*innen und Studierenden	193
Tabelle 5.22	Mittelwerte (Standardabweichungen) der Globalurteile pro Video und Statusgruppe	196
Tabelle 5.23	Einfaktorielle ANOVA der drei Erklärungen zum Thema Tonleiter mit <i>Bonferroni</i> -Korrektur	197
Tabelle 5.24	Unterschiede zwischen Erklärungen mit optischer Repräsentation und der Kombination aus optischer und akustischer Repräsentation unter Berücksichtigung der Statusgruppen	197
Tabelle 5.25	<i>Post-hoc</i> -Tests mit <i>Bonferroni</i> -Korrektur zu den Erklärungen mit akustischer Repräsentation und der Kombination aus optischer und akustischer Repräsentation	198

Tabelle 5.26	Unterschiede zwischen Erklärungen mit akustischer Repräsentation und der Kombination aus optischer und akustischer Repräsentation unter Berücksichtigung der Statusgruppen	199
Tabelle 5.27	Deskriptive Ergebnisse pro Statusgruppe zum Wunsch nach akustischer beziehungsweise optischer Repräsentation	199
Tabelle 5.28	Mittelwerte und Standardabweichungen über alle Studienteilnehmer*innen zu den fünf Merkmalen guten Erklärens, zur Persönlichkeitswirkung und zum Globalurteil	207
Tabelle 5.29	Mittelwerte und Standardabweichungen der Schüler*innen zu den fünf Merkmalen guten Erklärens, zur Persönlichkeitswirkung und zum Globalurteil	208
Tabelle 5.30	Mittelwerte und Standardabweichungen der Studierenden zu den fünf Merkmalen guten Erklärens, zur Persönlichkeitswirkung und zum Globalurteil	209
Tabelle 5.31	Mittelwerte und Standardabweichungen der Lehrenden zu den fünf Merkmalen guten Erklärens, zur Persönlichkeitswirkung und zum Globalurteil	210
Tabelle 5.32	Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für alle Studienteilnehmer*innen	211
Tabelle 5.33	Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für die Gruppe der Schüler*innen	211
Tabelle 5.34	Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für die Gruppe der Studierenden	212
Tabelle 5.35	Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für die Gruppe der Lehrenden	213
Tabelle 5.36	Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach Schüler*innen ($N_{max} = 148$) und den sieben Videos geordneten Datenstruktur	215
Tabelle 5.37	Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach Studierenden ($N_{max} = 61$) und den sieben Videos geordneten Datenstruktur	217
Tabelle 5.38	Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach Lehrenden ($N_{max} = 62$) und den sieben Videos geordneten Datenstruktur	219

Tabelle 5.39	Übersicht der drei hierarchischen linearen Regressionsmodelle getrennt nach Statusgruppe	220
Tabelle 5.40	Hierarchische lineare Regressionsmodelle mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach allen Studienteilnehmer*innen ($N_{max} = 271$) und den sieben Videos geordneten Datenstruktur	221
Tabelle 5.41	Anzahl der unterschiedlichen Scores sowie durchschnittlicher Score pro Erklärthema in den Vorwissenstests für alle Schüler*innen ($N = 142$)	244
Tabelle 5.42	Verteilung der Gesamtscores aller Schüler*innen ($N_{max} = 142$)	244
Tabelle 5.43	Hierarchisches lineares Regressionsmodell mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach Schüler*innen ($N_{max} = 142$) und den sieben Videos geordneten Datenstruktur	245

Literatur

- Abrami, P. C., Leventhal, L. & Perry, R. P. (1982). Educational seduction. *Review of Educational Research*, 52, 446–464. <https://doi.org/10.3102/00346543052003446>
- Aebli, H. (1983). *Zwölf Grundformen des Lehrens. Eine Allgemeine Didaktik auf psychologischer Grundlage*. Klett-Cotta.
- Aeschbacher, U. (2009). Eine Lanze fürs Erklären. *Beiträge zur Lehrerbildung*, 27(3), 431–437.
- Ahlers, M. & Seifert, A. (2015). Sprachliche Heterogenität im Musikunterricht. In A. Niessen & J. Knigge (Hrsg.), *Theoretische Rahmung und Theoriebildung in der musikpädagogischen Forschung* (S. 235–249). Waxmann.
- Allhoff, D.-W. & Allhoff, W. (2010). *Rhetorik und Kommunikation. Ein Lehr- und Übungsbuch* (15. Aufl.). Ernst Reinhardt Verlag.
- Alt, M. (1968). *Didaktik der Musik. Orientierung am Kunstwerk*. Pädagogischer Verlag Schwann.
- Andersen, J. F. (1979). Teacher Immediacy as a Predictor of Teaching Effectiveness. *Annals of the International Communication Association*, 3(1), 543–559. <https://doi.org/10.1080/23808985.1979.11923782>
- Ardila-Mantilla, N. (2018). Außerinstitutionelle Lernräume. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 397–405). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Aubussen, P. T., Harrison, A. G. & Ritchie, S. M. (2006). Metaphor and analogy. Serious thought in science education. In P. T. Aubussen, A. G. Harrison & S. M. Ritchie (Hrsg.), *Metaphor and Analogy in Science Education* (S. 1–9). Springer. https://doi.org/10.1007/1-4020-3830-5_1
- Babad, E. (2007). Teachers' nonverbal behavior and its effects on students. In R. Perry & J. Smart (Hrsg.), *The scholarship of teaching and learning: An evidence-based perspective* (S. 201–261). Springer. https://doi.org/10.1007/1-4020-5742-3_7
- Bailer, N. (2009). *Musikerziehung im Berufsverlauf. Eine empirische Studie über Musiklehrerinnen und Musiklehrer*. Universal Edition.
- Ball, D. L., Hill, H. H. & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, 29(3), 14–46.
- Bartelborth, T. (2007). *Erklären*. Verlag de Gruyter. <https://doi.org/10.1515/9783110916867>
- Barth, D. & Bubinger, A. (2015). Sprachbildung im Musikunterricht? Möglichkeiten und Perspektiven kultureller Teilhabe für mehrsprachige Kinder und Jugendliche. *Musikunterricht aktuell*, 2, 10–13.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A. & Krivitsky, P. N. (2022). lme4: Linear mixed-effects models using Eigen and S4 [Software]. <https://github.com/lme4/lme4/> (R package version 1.1-28).
- Baumert, J. (2002). Deutschland im internationalen Bildungsvergleich. In N. Killius, J. Kluge & L. Reisch (Hrsg.), *Die Zukunft der Bildung* (S. 100–150). Suhrkamp.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Wei, M. (2001). Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Leske + Budrich. https://doi.org/10.1007/978-3-322-83412-6_3
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaften*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>

- 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). Waxmann.
- Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (2014). *LehrplanPLUS Grundschule: Lehrplan für die bayerische Grundschule*. <https://www.lehrplanplus.bayern.de/schulart/grundschule> [07.08.2022].
- Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (2017a). *LehrplanPLUS Gymnasium: Lehrplan für das bayerische Gymnasium*. <https://www.lehrplanplus.bayern.de/schulart/gymnasium> [07.08.2022].
- Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (2017b). *LehrplanPLUS Realschule: Lehrplan für die bayerische Realschule*. <https://www.lehrplanplus.bayern.de/schulart/realschule> [07.08.2022]
- Becker-Mrotzek, M. (2004). *Schreibentwicklung und Textproduktion. Der Erwerb der Schreibfertigkeit am Beispiel der Bedienungsanleitung*. Verlag für Gesprächsforschung.
- Becker-Mrotzek, M., Schramm, K., Thürmann, E. & Vollmer, H. J. (2013). *Sprache im Fach. Sprachlichkeit und fachliches Lernen*. Waxmann.
- Beiderwieden, R. (2008). *Musik Unterrichten*. Bosse.
- Belgard, M., Rosenshine, B. & Gage, N. L. (1968). The teacher's effectiveness in Explaining: evidence on its generality and correlation with pupils' ratings and attention scores. In N. L. Gage (Hrsg.), *Explorations of the Teachers' Effectiveness in Explaining*. Stanford Center for Research and Development in Teaching, Stanford University (Technical Report No. 4).
- Benz, M. (2021). Der Musikunterricht sollte von einer Haltung getragen werden, mit der Unvorgeesehenes gewürdigt und produktiv genutzt wird. In V. Krupp (Hrsg.), *Wirksamer Musikunterricht*. Schneider Verlag Hohengehren.
- Berkemeier, A. (2009). „Wissen macht Ah!“ SchülerInnen erklären SchülerInnen Orthographieregeln medial. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (2. Aufl., S. 57–66). Stauffenburg Linguistik.
- 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>
- Biegholdt, G. (2013). *Musik unterrichten. Grundlagen – Gestaltung – Auswertung*. Kallmeyer.
- Bieneck-Hempel, M. (2009). *Musiktheorie im Musikunterricht an niedersächsischen Gymnasien. Eine Bestandsaufnahme mit statistischer Auswertung*. Die Blaue Eule.
- Bliese, P., Chen, G., Downes, P., Schepker, D. & Lang, J. (2022). multilevel: Multilevel Functions. [Software]. <https://www.r-project.org> (R package version 2.7) [07.08.2022].
- Blömeke, S., Bremerich-Vos, A., Haudeck, H., Kaiser, G., Nold, G., Schwippert, K. & Willenberg, H. (Hrsg.) (2011). *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT*. Waxmann.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. Competence Viewed as a Continuum. *Zeitschrift für Psychologie*, 223(1), S. 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Blömeke, S., Kaiser, G. & Lehmann, R. (2010a). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*. Waxmann.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.) (2010b). *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Waxmann.
- Blömeke, S., König, J., Suhl, U., Hoth, J. & Döhrmann, M. (2015). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests. *Zeitschrift für Pädagogik*, 61(3), S. 310–327. <https://doi.org/10.25656/01:15350>

- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-12770-0>
- Bose, I., Hirschfeld, U., Neuber, B. & Stock, E. (2016). *Einführung in die Sprechwissenschaft*. Narr Verlag.
- Bossen, A. (2019). *Sprachbewusster Musikunterricht. Problematisierung sprachdidaktischer Ansätze und Perspektiven einer Sprachbildung im Fach*. Waxmann.
- Bossen, A. & Jank, B. (2017). *Sprache im Musikunterricht. Ausgewählte Aspekte sprachbewussten Handelns im Kontext von Inklusion*. Universitätsverlag Potsdam.
- Brandstätter, U. (2010). Auf der Suche nach dem Licht didaktischer Theorien: Die Musikstunden im Vergleich. In C. Wallbaum (Hrsg.), *Perspektiven der Musikdidaktik. Drei Schulstunden im Licht der Theorien* (S. 29–46). Olms.
- Bredel, U. & Maaß, C. (2016). *Leichte Sprache. Theoretische Grundlagen – Orientierung für die Praxis*. Bibliographisches Institut.
- Brenk, M. (2014). Nur noch Etüden? – Kritisch-konstruktive Anmerkungen zur Kompetenzorientierung im Musikunterricht. *Zeitschrift für Kritische Musikpädagogik*, 2014, 57–67. <http://www.zfkm.org/14-brenk.pdf>
- Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens*. Hans Huber.
- Brophy, J. (2000). *Teaching*. International Academy of Education. http://www.ibe.unesco.org/fileadmin/user_upload/archive/Publications/educationalpracticesseriespdf/prac01e.pdf [04.08.2022].
- Brown, G. (2006). Explaining. In O. Hargie (Hrsg.), *The handbook of communication skills* (3. Aufl., S. 195–228). Routledge.
- Brown, G. & Hatton, N. (1982). *Explanations and Explaining. A teaching skills workbook*. Macmillan Education. <https://doi.org/10.3109/01421598209034743>
- Bruckmaier, G., Krauss, S., Blum, W. & Leiss, D. (2016). Measuring mathematics teachers' professional competence by using video clips (COACTIV video). *ZDM Mathematics Education*, 48, 111–124. <https://doi.org/10.1007/s11858-016-0772-1>
- Bruhn, H. (2019). Notation als mediale Darstellung von Musik. In H. Schramm (Hrsg.), *Handbuch Musik und Medien. Interdisziplinärer Überblick über die Mediengeschichte der Musik* (2. Aufl., S. 255–271). Springer. https://doi.org/10.1007/978-3-658-21899-7_10
- Brunner, E. (2017). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik*, 39, 257–284. <https://doi.org/10.1007/s13138-017-0122-z>
- Buchholtz, N., Kaiser, G. & Stancel-Piątak, A. (2011). Professionelles Wissen von Studierenden des Lehramts Mathematik. In S. Blömeke, A. Bremerich-Vos, H. Haudeck, G. Kaiser, G. Nold, K. Schwippert & H. Willenberg (Hrsg.), *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen: Erste Ergebnisse aus TEDS-LT* (S. 101–134). Waxmann.
- Burger, T. (2018). *Rhetorik für Lehrkräfte*. Klinkhardt/UTB. <https://doi.org/10.36198/9783838550466>
- Busch, T. & Kranefeld, U. (2013). Individuelle Förderung im instrumentalen Gruppenunterricht. Ein Aspekt von Lehrendenkompetenz im Programm „Jedem Kind ein Instrument“. In A. Lehmann-Wermser & M. Krause-Benz (Hrsg.), *Musiklehrer(-bildung) im Fokus musikpädagogischer Forschung* (S. 99–116). Waxmann.
- Buschendorff, F. (2010). *200 Methoden für den Musikunterricht. Praxisorientierte Ideen für die Sekundarstufe*. Verlag an der Ruhr.
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* (4. korr. und erw. Aufl.). Pearson.
- Bühner, M. & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler. Grundlagen und Umsetzung mit SPSS und R* (2. Aufl.). Pearson.

- Charalambous, C. Y., Hill, H. C. & Ball, D. L. (2011). Prospective teachers' learning to provide instructional explanations: How does it look and what might it take? *Journal of Mathematics Teacher Education*, 14(6), 441–463. <https://doi.org/10.1007/s10857-011-9182-z>
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Waxmann.
- Coenen, H. G. (2002). *Analogie und Metapher*. De Gruyter. <https://doi.org/10.1515/9783110894639>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.) Erlbaum.
- Cramer, C., König, J., Rothland, M. & Blömeke, S. (2020). *Handbuch Lehrerinnen- und Lehrerbildung*. Klinkhardt/UTB. <https://doi.org/10.36198/9783838554730>
- Cramer, C. & Schreiber, F. (2018). Subject Didactics and Educational Sciences. Relationships and Their Implications for Teacher Education from the Viewpoint of Educational Sciences. *Research in Subject-matter Teaching and Learning*, 1(2), 150–164. <https://doi.org/10.23770/rt1818>
- Cvetko, A. J. (2015). *Geschichten erzählen als Methode im Musikunterricht. Historische und empirische Studien*. LIT.
- Dartsch, M. (2014). *Musik lernen und unterrichten. Eine Einführung in die Musikpädagogik*. Breitkopf & Härtel.
- Dartsch, M. (2018). Lernformen und Lernwege. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 169–176). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Dartsch, M., Knigge, J., Niessen, A., Platz, F. & Stöger, C. (Hrsg.) (2018). *Handbuch Musikpädagogik: Grundlagen – Forschung – Diskurse*. Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Dartsch, M. & Platz, F. (2018). Perspektiven auf Lernen. Einleitung. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 157–158). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Decristan, J., Hess, M., Holzberger, D. & Praetorius, A.-K. (2020). Oberflächen- und Tiefenmerkmale. Eine Reflexion zweier prominenter Begriffe der Unterrichtsforschung. *Zeitschrift für Pädagogik*, 66(Beiheft 1/20), 102–116.
- DESI-Konsortium (Hrsg.) (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Beltz.
- Detterbeck, M. & Schmidt-Oberländer, G. (2019). *MusiX 1. Das Kursbuch Musik für den Unterricht an allgemeinbildenden Schulen*. Helbling
- Diederich, J. & Tenorth, H.-E. (1997). *Theorie der Schule. Ein Studienbuch zur Geschichte, Funktion und Gestaltung*. Cornelsen Scriptor.
- Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. *Zeitschrift für Pädagogik*, 41(Beiheft), 73–92.
- Dorfman, J. (2013). *Theory and practice of technology-based music instruction*. Oxford University Press. <https://doi.org/10.1093/oso/9780199795581.001.0001>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial und Humanwissenschaften* (5. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Dressler, B. (2007). Modi der Weltbegegnung als Gegenstand fachdidaktischer Analysen. *Journal für Mathematik-Didaktik*, 28, 249–262. <https://doi.org/10.1007/BF03339348>
- Dreßler, S. (2016) Ästhetische Erfahrung im Musikunterricht oder Begegnung mit Klingendem, die uns „aus dem Anzug stößt“. In A. Lehmann-Wermser (Hrsg.), *Musikdidaktische Konzeptionen. Ein Studienbuch* (S. 45–64). Wißner.

- Dreßler, S. (2017). Musiklernen im problemorientierten Musikunterricht. In B. Hofmann (Hrsg.), *Plan | mä | ßig. Schulmusik unter den Vorzeichen von Bildungsstandards und Kompetenzorientierung: Beiträge zu den Tagen der Bayerischen Schulmusik 2016* (S. 45–58). Helbling.
- Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit. Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht*. Waxmann.
- Dudenredaktion. (o. J.). Erklären. In *Duden online*. <https://www.duden.de/rechtschreibung/erklaren> [07.08.2022].
- Dyllick, N. (2018). Singen in der Schule. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 299–302). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Eberhard, D. M. (2010). *Ursachen von Unterrichtsstörungen im Fach Musik aus Sicht der Beteiligten und Entwurf eines Diagnosebogens zur Metakommunikation im Musikunterricht der Bayerischen Realschule*. [Dissertation]. <https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/1430> [07.08.2022].
- Eberhard, D. M. (2016). *Musik unterrichten: planen, durchführen, reflektieren. Sekundarstufe I + II*. Cornelsen.
- Ehlich, K. (2009). Erklären verstehen – Erklären und Verstehen. In Vogt, R. (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven*. (2. Aufl., S. 11–24). Stauffenberg.
- Ehninger, J., Knigge, J., Schurig, M. & Rolle, C. (2021). A New Measurement Instrument for Music-Related Argumentative Competence: The MARKO Competency Test and Competency Model. *Frontiers in Education*, 6. <https://doi.org/10.3389/educ.2021.668538>
- Ehras, C., Asen-Molz, K., Frei, M., Schilcher, A. & Krauss, S. (2021). Erklären lernen – ein Seminarkonzept zur Förderung von Erklärkompetenz durch Videografie als Reflexionsanlass. In E. Matthes, S. T. Siegel & T. Heiland (Hrsg.), *Lehrvideos – das Bildungsmedium der Zukunft? Erziehungswissenschaftliche und fachdidaktische Perspektiven* (S. 203–212). Klinkhardt.
- Ehrenforth, K. H. (1971). *Verstehen und Auslegen: die hermeneutischen Grundlagen einer Lehre von der didaktischen Interpretation der Musik*. Diesterweg.
- Ehrenforth, K. H. (1993). Musik als Leben. Zu einer lebensweltlich orientierten ästhetischen Hermeneutik. *Musik und Bildung*, 6, 14–19.
- Ehrenforth, K. H. (2010). *Geschichte der musikalischen Bildung: Eine Kultur-, Sozial- und Ideengeschichte in 40 Stationen; von der antiken Hochkultur bis zur Gegenwart*. Schott.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (5. korr. Aufl.). Beltz.
- Ernst, A. (1991). *Lehren und Lernen im Instrumentalunterricht*. Schott.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive. Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für pädagogische Psychologie*, 28(3), 127–137. <https://doi.org/10.1024/1010-0652/a000129>
- Fauth, B., Göllner, R., Lenske, G., Praetorius, A.-K. & Wagner, W. (2020). Who Sees What? Conceptual Considerations on the Measurement of Teaching Quality from Different Perspectives. *Zeitschrift für Pädagogik*, 66(Beiheft 1/20), 138–155.
- Ferguson, R. & Danielson, C. (2014). How framework for teaching and Tripod 7Cs evidence distinguish key components of effective teaching. In T. Kane, K. Kerr & R. Pianta (Hrsg.), *Designing teacher evaluation systems* (S. 98–143). Jossey-Bass. <https://doi.org/10.1002/9781119210856.ch4>
- Fiedler, D. & Handschick, M. (2014). *Produktive Methoden im Test – Zum Stellenwert und zur Attraktivität produktiver Methoden im Musikunterricht an allgemeinbildenden Schulen*. PH Freiburg. <http://nbn-resolving.de/urn:nbn:de:bsz:frei129-opus-4409> [07.08.2022].

- Fiedler, D. & Müllensiefen, D. (2016). Struktur und Entwicklung von Musikalischem Selbstkonzept, Musikalischer Erfahrungheit und Interesse am Schulfach Musik: Eine empirische Längsschnittuntersuchung von Schülerinnen und Schülern (9 bis 17 Jahre) an Haupt-, Gemeinschafts- und Realschulen sowie Gymnasien in Baden-Württemberg. In J. Knigge & A. Niessen (Hrsg.), *Musikpädagogik und Erziehungswissenschaft* (S. 209–230). Waxmann.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4. Aufl.). SAGE Publications.
- Findeisen, S. (2017). *Fachdidaktische Kompetenzen angehender Lehrpersonen. Eine Untersuchung zum Erklären im Rechnungswesen*. Springer. <https://doi.org/10.1007/978-3-658-18390-5>
- Flämig, M. (1998). *Verstehen – Hören – Handeln: Destruktion und Rekonstruktion der Begriffe*. Wißner.
- Flick, U. (2009). *Sozialforschung: Methoden und Anwendungen. Ein Überblick für die BA-Studiengänge*. Rowohlt Verlag.
- Flick, U. (2011). *Triangulation. Eine Einführung* (3. Aufl.). VS Verlag für Sozialwissenschaften.
- Forge, S. (2016). *Die Nachhaltigkeit von Lehrerbildung am Beispiel des Singförderprojekts „SMS – Singen mach Sinn“*. [Dissertation]. <https://doi.org/10.17619/UNIPB/1-169>
- Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Krivitsky, P., Laboisiere, R., Maechler, M., Monette, G., Murdoch, D., Nilsson, H., Ogle, D., Phipley, B., ... Zeileis, A. (2022). car: Companion to Applied Regression. [Software]. <https://CRAN.R-project.org/package=car> (R package version 3.1-0).
- Raefel, U. & Scheidig, F. (2018). Mit Pragmatik zu professioneller Praxis? Der Core-Practices-Ansatz in der Lehrpersonenbildung. *Beiträge zur Lehrerinnen-und Lehrerbildung*, 36(3), 344–364. <https://doi.org/10.25656/01:18855>
- Frei, M. (2019). Probing als unterstützende Methode bei der Entwicklung eines Fragebogens. Ein Anwendungsbericht aus dem Forschungsprojekt FALKE-Musik. In G. Puffer, A. Becker, F. Körndle & K. Sprau (Hrsg.), *Musik – Pädagogik – Professionalität. Festschrift für Bernhard Hofmann zum 60. Geburtstag* (S. 121–136). Helbling.
- Frei, M., Asen-Molz, K., Hilbert, S., Schilcher, A. & Krauss, S. (2020). Die Wirksamkeit von Erklärvideos im Rahmen der Methode Flipped Classroom. In K. Kaspar, M. Becker-Mrotzek, S. Hofhues, J. König & D. Schmeinck (Hrsg.), *Bildung, Schule, Digitalisierung* (S. 284–290). Waxmann.
- Frei, M., Puffer, G., Hilbert, S. & Hofmann, B. (2022). Zwischen Generik und Spezifik – zur Konzeptualisierung von „gutem Erklären“ im schulischen Musikunterricht. In M. Göllner, J. Knigge A. Niessen & V. Weidner (Hrsg.), *43. Jahresband des Arbeitskreises Musikpädagogische Forschung [43rd Yearbook of the German Association for Research in Music Education]*. Waxmann.
- Frei, M., Puffer, G. & Hofmann, B. (i. Vorb.). Allgemeine Musiklehre – „gut“ erklärt. In A. Schilcher, S. Krauss, A. Lindl & S. Hilbert (Hrsg.), *Fachspezifische Lehrerkompetenzen im Erklären*. Beltz.
- Frey, A. (2014). Kompetenzmodelle und Standards in Lehrerbildung und Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. und erw. Aufl., S. 712–744). Waxmann.
- Fricke, M. (2017). FALCO-R: Professionswissen von Religionslehrkräften. Entwicklung eines Messinstruments zur fachspezifischen Lehrerkompetenz. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALCO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 291–335). Waxmann.

- Fröhlich-Schweiger, M. (2019). Singen, Sprechen, Präsentieren. Kreatives Spiel mit und Reflexion über Formen verbaler und nonverbaler Kommunikation. In G. Puffer, A. Becker, F. Körndle & K. Sprau (Hrsg.), *Musik – Pädagogik – Professionalität. Festschrift für Bernhard Hofmann zum 60. Geburtstag* (S. 293–302). Helbling.
- Fuchs, M. (2015). Ziele des Musikunterrichts. In M. Fuchs (Hrsg.), *Musikdidaktik Grundschule. Theoretische Grundlagen und Praxisvorschläge* (S. 88–103). Helbling.
- Fuchs, M. & Brunner, G. (2015). Handeln und begreifen: Aufbau musikalischer Fähigkeiten. In M. Fuchs (Hrsg.), *Musikdidaktik Grundschule. Theoretische Grundlagen und Praxisvorschläge* (S. 218–252). Helbling.
- Fuß, H.-U. (1997). Musik begreifen 1. Musiktheorie, Musiklehre. In S. Helms, R. Schneider & R. Weber (Hrsg.), *Handbuch Musikunterricht. Sekundarstufe I* (S. 153–178). Bosse.
- Gaier, L. (2020). *Gut erklärt? FALKE-D: Eine empirische Studie zu Erklärungen im Deutschunterricht aus unterschiedlicher Beobachterperspektiven*. Universität Regensburg Universitätsbibliothek. <https://epub.uni-regensburg.de/51210/> [07.08.2022].
- Gaul, M. (2009). *Musikunterricht aus Schülersicht*. Schott.
- Gautschi, P. (2009). *Guter Geschichtsunterricht: Grundlagen, Erkenntnisse, Hinweise*. Wochenschau Verlag.
- Gebauer, H. (2013). „Beschreibt doch mal die Form, die wir gerade gemacht haben.“ Kognitive Aktivierung im Musikunterricht. In A. Lehmann-Wermser & M. Krause-Benz (Hrsg.), *Musiklehrer(-bildung) im Fokus musikpädagogischer Forschung* (S. 61–79). Waxmann.
- Gebauer, H. (2016). *Kognitive Aktivierung im Musikunterricht: Eine qualitative Videostudie*. Dissertation. LIT.
- Geelan, D. (2012). Teacher Explanation. In B. Fraser, K. Tobin & C. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 987–999). Springer. https://doi.org/10.1007/978-1-4020-9041-7_65
- Geelan, D. (2013). Teacher Explanation of Physics Concepts: A Video Study. *Research in Science Education*, 43, 1751–1762. <https://doi.org/10.1007/s11165-012-9336-8>
- Geiser, C. & Eid, M. (2006). Profilanalyse. In F. Petermann & M. Eid (Hrsg.), *Handbuch der Psychologischen Diagnostik* (S. 318–313). Hogrefe.
- Geuen, H. (2018). Musikbegriffe. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 18–24). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Gindele, V. & Voss, T. (2017). Pädagogisch-psychologisches Wissen: Zusammenhang mit Indikatoren des beruflichen Erfolgs angehender Lehrkräfte. *Zeitschrift für Bildungsforschung*, 7(3), 255–272. <https://doi.org/10.1007/s35834-017-0192-5>
- Gläser-Zikuda, M., Seidel, T., Rohlf, C., Gröschner, A. & Ziegelbauer, S. (2012). Mixed Methods in der empirischen Bildungsforschung. Eine Einführung in die Thematik. In M. Gläser-Zikuda, T. Seidel, C. Rohlf, A. Gröschner & S. Ziegelbauer (Hrsg.), *Mixed Methods in der empirischen Bildungsforschung* (S. 7–13). Waxmann.
- Godau, M. (2018). Wie kommen die *Dinge* in den Musikunterricht? Zur Materialität musikpädagogischer Praxis am Beispiel divergierender Orientierungen im Kontext unterrichtsbezogenen Handelns angehender Lehrkräfte. In B. Clausen & S. Dreßler (Hrsg.), *Soziale Aspekte des Musiklernens. Social aspects of music learning* (S. 43–55). Waxmann.
- Godau, M. & Fiedler, D. (2018) Erfassung des Professionswissens von Musiklehrkräften: Validierung einer deutschen Übersetzung eines Selbstauskunftsfragebogens zur Erfassung des Musical Technological Pedagogical And Content Knowledge (MTPACK) In B. Clausen & S. Dreßler (Hrsg.), *Soziale Aspekte des Musiklernens. Social aspects of music learning* (S. 185–205). Waxmann

- Göllner, M. (2017). *Perspektiven von Lehrenden und SchülerInnen auf Bläserklassenunterricht: Eine qualitative Interviewstudie*. Waxmann.
- 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*. (S. 122–136). Waxmann
- Göllner, R., Wagner, W., Klieme, E., Lüdtke, O., Nagengast, B. & Trautwein, U. (2016). Erfassung der Unterrichtsqualität mithilfe von Schülerurteilen: Chancen, Grenzen und Forschungsperspektiven. In Bundesministerium für Bildung und Forschung (Hrsg.), *Forschungsvorhaben in Anknüpfung an Large-Scale-Assessments* (S. 63–82). Bertelsmann Verlag.
- Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96(3), 606–633. <https://doi.org/10.1525/aa.1994.96.3.02a00100>
- Grimmer, F. & Lessing, W. (2008). *Künstler als Pädagogen. Grundlagen und Bedingungen einer verantwortungsvollen Instrumental Didaktik*. Schott.
- Groeben, N., Wahl, D., Scheele, B. & Schlee, J. (1988). *Das Forschungsprogramm Subjektive Theorien. Eine Einführung in die Psychologie des reflexiven Subjekts*. Francke.
- Grossman, P. L., Hammerness, K. & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching*, 15(2), 273–289. <https://doi.org/10.1080/13540600902875340>
- Gruber, H. & Stöger, H. (2011). Experten-Novizen-Paradigma. In E. Kiel & K. Zierer (Hrsg.), *Unterrichtsgestaltung als Gegenstand der Wissenschaft* (S. 247–264). Schneider Verlag Hohengehren.
- Gruhn, W. (1994). Musiklernen – Der Aufbau musikalischer Repräsentationen. In G. Olias (Hrsg.), *Musiklernen. Aneignung des Unbekannten* (S. 9–31). Die Blaue Eule.
- Gruhn, W. (2003a). *Geschichte der Musikerziehung: Eine Kultur- und Sozialgeschichte vom Gesangsunterricht der Aufklärungspädagogik zu ästhetisch-kultureller Bildung*. Wolke.
- Gruhn, W. (2003b). *Lernziel Musik*. Olms.
- Grüner, G. (1967). *Die didaktische Reduktion als Kernstück der Didaktik*. Waxmann.
- Gunga, E., Schilcher, A., Kranich, W. & Gegner, C. (i. Vorb.). FALKE-q-SE: Einfluss des Sprech- und Körperausdrucks von Lehrkräften auf die wahrgenommene Qualität von schulischen Erklärungen. In A. Schilcher, A. Lindl, S. Krauss & S. Hilbert (Hrsg.), *Fachspezifische Lehrerkompetenzen im Erklären*. Beltz Juventa.
- Haas, M., Nonte, S., Krieg, M. & Stubbe, T. C. (2019). Unterrichtsqualität in Musikklassen. Befunde aus der quasi-experimentellen Studie ProBiNi. In V. Weidner & C. Rolle (Hrsg.), *Praxen und Diskurse aus Sicht musikpädagogischer Forschung* (S. 137–154). Waxmann.
- Hafen, R., Hartog, T., Kehler, U., Kleine-Huster, D. & Bohn, A. (2011). Musikunterricht auf dem Prüfstand – ästhetische Erfahrung im Fokus der Empirie. In N. Logemann (Hrsg.), *Unterrichtsqualität und fachdidaktische Forschung. Modelle und Instrumente zur Messung fachspezifischer Lernbedingungen und Kompetenzen* (S. 187–220). Waxmann.
- Hammel, L. (2011). *Selbstkonzepte fachfremd unterrichtender Musiklehrerinnen und Musiklehrer an Grundschulen: Eine Grounded-Theory-Studie*. LIT.
- Hargie, O. (2013). *Die Kunst der Kommunikation. Forschung, Theorie, Praxis*. Hans Huber.
- Harnischmacher, C. (2002). Verbessert Sehen das Hören? Eine experimentelle Studie zum Einfluss der Nutzung von computergestützter Visualisierung auf die Melodiewahrnehmung und das Benennen von Tonhöhenrichtungen. In R.-D. Kraemer (Hrsg.), *Multimedia als Gegenstand musikpädagogischer Forschung* (S. 97–113). Die Blaue Eule.
- Harnischmacher, C. (2018). Motivation. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 221–228). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>

- Harnischmacher, C., Blum, K. & Hofbauer, V. C. (2017). Kompetenzorientierung und Motivation von Lehrern im Musikunterricht. In A. J. Cvetko & C. Rolle (Hrsg.), *Musikpädagogik und Kulturwissenschaft* (S. 269–288). Waxmann.
- Harnischmacher, C. & Hofbauer, V. C. (2011). Wahrnehmungsdimensionen des Musikunterrichts: Eine explorative Studie zur Unterrichtsbeobachtung von Schülern, Studenten, Lehramtsanwärtern und Lehrern. *Beiträge empirischer Musikpädagogik (b:em)*, 2(2). <https://www.b-em.info/index.php/ojs/article/view/60> [07.08.2022].
- Harnischmacher, C., Hofbauer, V. C. & Schulz-Heidorf, K. (2019). Warum Musik wählen? Eine mehrbenenanalytische Studie zur Vorhersage der Wahlbereitschaft zum Fach Musik durch die Motivation der Schüler*innen und der individuellen Förderung von Musiklehrkräften im Musikunterricht. In V. Weidner & C. Rolle (Hrsg.), *Praxen und Diskurse aus Sicht musikpädagogischer Forschung* (S. 189–204). Waxmann.
- Harrison, A. G. & Treagust, D. F. (1993). Teaching with Analogies: A Case Study in Grade-10 Optics. *Journal of research in science teaching*, 30(10), 1291–1307. <https://doi.org/10.1002/tea.3660301010>
- Hartig, K. (2015). *Der Einfluss der Präsentationsgeschwindigkeit auf das Wahrnehmen und Verstehen beim Lernen mit einer Animation* [Dissertation]. Universität Koblenz-Landau Universitätsbibliothek.
- Hasselhorn, J. (2015). *Messbarkeit musikpraktischer Kompetenzen von Schülerinnen und Schülern: Entwicklung und empirische Validierung eines Kompetenzmodells*. Waxmann.
- Hasselhorn, J. & Knigge, J. (2018). Kompetenz und Expertise. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 197–207). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Hasselhorn, J. & Lehmann, A. C. (2014). Entwicklung eines empirisch überprüfbareren Modells musikpraktischer Kompetenz (KOPRA-M). In B. Clausen (Hrsg.), *Teilhabe und Gerechtigkeit/ Participation and Equity* (S. 77–93). Waxmann.
- Hasselhorn, J. & Lehmann, A. C. (2015). Leistungsheterogenität im Musikunterricht. Eine empirische Untersuchung zu Leistungsunterschieden im Bereich der Musikpraxis in Jahrgangsstufe 9. In A. Niessen & J. Knigge (Hrsg.), *Theoretische Rahmung und Theoriebildung in der musikpädagogischen Forschung* (S. 199–219). Waxmann.
- Hattie, J., Beywl, W. & Zierer, K. (2014). *Lernen sichtbar machen* (2., korr. Aufl.). Schneider-Verlag.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Heilmann, C. M. (2002). *Interventionen im Gespräch. Neue Ansätze der Sprechwissenschaft*. Niemeyer. <https://doi.org/10.1515/9783110918618>
- Hellberg, B. (2018). Zwischen klingenden Rohdaten und sprachlicher Transformation. In C. Moritz & M. Corsten (Hrsg.), *Handbuch Qualitative Videoanalyse* (S. 217–234). Springer. https://doi.org/10.1007/978-3-658-15894-1_13
- Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts; Franz Emanuel Weinert gewidmet* (7. Aufl., [aktualisierte Aufl., berücksichtigt die Hattie-Studien]). Klett.
- Helmke, A. & Schrader, W. (2008). Merkmale der Unterrichtsqualität: Potenzial, Reichweite und Grenzen. *SEMINAR – Lehrerbildung und Schule*, 3, 17–47.
- Helsper, W. (2014). Lehrerprofessionalität – der strukturtheoretische Professionsansatz zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. und erw. Aufl., S. 216–240). Waxmann.

- Hempel, C. (2008). Visualisierung von Musik. In S. Weinacht & H. Scherer (Hrsg.), *Wissenschaftliche Perspektiven auf Musik und Medien* (S. 193–204). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90859-5_12
- Hempel, C. G. & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2), 135–175. <https://doi.org/10.1086/286983>
- Herrmann, C. & Gerlach, E. (2020). Unterrichtsqualität im Fach Sport. Ein Überblicksbeitrag zum Forschungsstand in Theorie und Empirie. *Unterrichtswissenschaft*, 48, 361–384. <https://doi.org/10.1007/s42010-020-00080-w>
- Heß, C. (2018a). Klassenmusizieren in der Schule. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 306–308). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Heß, F. (2011). Musikunterricht zwischen Sach- und Fachinteresse: Ergebnisse aus der Pilotstudie ‚Musikunterricht aus Schülersicht‘. *Beiträge empirischer Musikpädagogik/ Bulletin of empirical music education research (b:em)*, 2(1). <https://b-em.info/index.php/ojs/article/view/44> [07.08.2022].
- Heß, F. (2013). „Klassik“ und Musikgeschichte im Unterricht. In W. Jank (Hrsg.), *MusikDidaktik: Praxishandbuch für die Sekundarstufe I und II* (5. Aufl., S. 204–212). Cornelsen Scriptor.
- Heß, F. (2018b). *Gendersensibler Musikunterricht. Empirische Studien und didaktische Konsequenzen*. Springer VS. <https://doi.org/10.1007/978-3-658-19166-5>
- Hilbert, S., Stadler, M., Lindl, A., Naumann, F. & Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *TPM*, 26(1), 101–119. <https://doi.org/10.4473/TPM26.1.6>.
- Hill, H. C., Charalambous, C. Y. & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56–64. <https://doi.org/10.3102/0013189X12437203>
- Hofbauer, V. C. (2017). *Motivation von Musiklehrern. Zum Einfluss der Motivation und Expertise auf die Stressbewältigung*. Springer. <https://doi.org/10.1007/978-3-658-15206-2>
- Hofmann, B. (2011). Experten für Musiklehre? Auf der Suche nach der „guten“ Musiklehrkraft. In M. D. Loritz, A. Becker & A. M. Eberhard (Hrsg.), *Musik – Pädagogisch – Gedacht. Reflexionen, Forschungs- und Praxisfelder; Festschrift für Rudolf-Dieter Kraemer zum 65. Geburtstag* (S. 95–103). Wißner.
- Hohenstein, C. (2006). *Erklärendes Handeln im Wissenschaftlichen Vortrag. Ein Vergleich des Deutschen mit dem Japanischen*. Iudicium.
- Hohenstein, C. (2009). Interkulturelle Aspekte des Erklärens. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (2. Aufl., S. 37–56). Stauffenberg.
- Hörmann, S. & Meidel, E. (2016). Orientierung im Begriffsdschungel – terminologische und fachstrukturelle Perspektiven zur Profilierung der Musikpädagogik und Musikdidaktik. In B. Clausen, A. Cvetko, S. Hörmann, M. Krause-Benz & S. Kruse-Weber (Hrsg.), *Grundlagentexte Wissenschaftlicher Musikpädagogik. Begriffe, Positionen, Perspektiven im systematischen Fokus* (S. 11–68). Waxmann.
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M., McElvany, N., Stubbe, T. C. & Valtin, R. (Hrsg.) (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann.
- Hußmann, S., Kranefeld, U., Kuhl, J. & Schlebrowski, D. (2018). Das geschachtelte Tetraeder und inklusionsorientierte Designprinzipien als Modell für Entwicklung und Forschung in einer inklusionsorientierten Lehrerinnen- und Lehrerbildung. In S. Hußmann & B. Welzel (Hrsg.), *DoProfil – Das Dortmunder Profil für inklusionsorientierte Lehrerinnen- und Lehrerbildung* (S. 11–26). Waxmann.

- Jank, W. (Hrsg.) (2013). *MusikDidaktik: Praxishandbuch für die Sekundarstufe I und II* (5., überarb. Neuauf.). Cornelsen Scriptor.
- Jank, W. (2018). Unterrichtsgestaltung: schulischer Musikunterricht. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 355–360). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Jank, W., Knigge, J. & Niessen, A. (2020). Musikdidaktik – Bestandsaufnahme und Forschungsperspektiven. In M. Rothgangel, U. Abraham, H. Bayrhuber, V. Frederking, W. Jank & H. Vollmer (Hrsg.), *Lernen im Fach und über das Fach hinaus. Bestandsaufnahmen und Forschungsperspektiven aus 17 Fachdidaktiken im Vergleich* (S. 262–288). Waxmann.
- Jank, W. & Meyer, H. (2013). Zur Unterrichtsplanung. In W. Jank (Hrsg.), *MusikDidaktik. Praxishandbuch für die Sekundarstufe I und II* (5., überarb. Neuauf., S. 132–140). Cornelsen Scriptor.
- Jank, W., Meyer, H. & Ott, T. (1986) Zur Person des Lehrers im Musikunterricht: Methodische Probleme und Perspektiven zu einem Konzept offenen Musikunterrichts. In H. J. Kaiser (Hrsg.), *Unterrichtsforschung* (S. 87–131). Laaber Verlag.
- Jank, W. & Schmidt-Oberländer, G. (2010). *Music Step by Step. Aufbauender Musikunterricht in der Sekundarstufe I. Lehrerhandbuch*. Helbling.
- Jeismann, A.-L. & Kranefeld, U. (2021). (Un-)Eindeutige Anregungen. Zur Rekonstruktion von Handlungsmustern bei der Begleitung von Prozessen des Musik-Erfindens. In J. Hasselhorn, O. Kautny & F. Platz (Hrsg.), *Musikpädagogik im Spannungsfeld von Reflexion und Intervention* (S. 135–153). Waxmann.
- Jordan, A.-K., Knigge, J., Lehmann, A. C., Niessen, A. & Lehmann-Wermser, A. (2012). Entwicklung und Validierung eines Kompetenzmodells im Fach Musik – Wahrnehmen und Kontextualisieren von Musik. *Zeitschrift für Pädagogik*, 58(4), 500–521.
- Jörissen, B. (2015). Bildung der Dinge: Design und Subjektivierung. In B. Jörissen & T. Meyer (Hrsg.), *Subjekt Medium Bildung* (S. 215–234). Springer. https://doi.org/10.1007/978-3-658-06171-5_11
- Jünger, H. (2006) Musikunterricht ohne Schulbuch? Empirische Untersuchung zur Verwendung von Musiklehrbüchern In N. Knolle (Hrsg.), *Lehr- und Lernforschung in der Musikpädagogik* (S. 201–221). Die Blaue Eule.
- Kaiser, H. J. (2002). Musik in der Schule? – Musik in der Schule! Lernprozesse als ästhetische Bildungspraxis. *Zeitschrift für Kritische Musikpädagogik, Sonderedition 1*. https://www.zfkm.org/sonder02-kaiser_a.pdf [07.08.2022].
- Kaiser, H. J. (2010). Verständige Musikpraxis. Eine Antwort auf Legitimationsdefizite des Klassenmusizierens. *Zeitschrift für Kritische Musikpädagogik*, 47–68. <http://zfkm.org/10-kaiser.pdf> [07.08.2022].
- Kaiser, H. J. & Nolte, E. (1989) *Musikdidaktik: Sachverhalte – Argumente – Begründungen*. Schott.
- Keller-Schneider, M. & Hericks, U. (2014). Forschungen zum Berufseinstieg. Übergang von der Ausbildung in den Beruf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. und erw. Aufl., S. 386–407). Waxmann.
- Kemmelmeyer, K.-J. & Nykrin, R. (2004). *Spielpläne Musik 5/6* (Neue Ausgabe C Bayern) – Schülerbuch. Klett.
- Kiel, E. (1999). *Erklären als didaktisches Handeln*. Ergon.
- Kiel, E. (2014). Forschung zum Lehrerhandeln. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. und erw. Aufl., S. 937–943). Waxmann.
- Kim, K. Y. & Crowley, K. (2010). Negotiating the Goal of Museum Inquiry: How Families Engineer and Experiment. In M. K. Stein & L. Kucan (Hrsg.), *Instructional Explanations in the Disciplines* (S. 51–65). Springer. https://doi.org/10.1007/978-1-4419-0594-9_5

- Kirchhoff, P. (2017). FALKO-E: Fachspezifisches professionelles Wissen von Englischlehrkräften. Entwicklung und Validierung eines domänenspezifischen Testinstruments. In S. Krauss, A. Lindl & A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 113–152). Waxmann.
- Kirschner, P. A., Verschaffel, L., Star, J. & Van Dooren, W. (2017). There is more variation within than across domains: an interview with Paul A. Kirschner about applying cognitive psychology based instructional design principles in mathematics teaching and learning. *ZDM Mathematics Education*, 49, 637–643. <https://doi.org/10.1007/s11858-017-0875-3>
- Kleickmann, T., Praetorius, A.-K. & Riecke-Baulecke, T. (2019). Beurteilung von Unterrichtsqualität. In F. Zimmermann, J. Möller & T. Riecke-Baulecke (Hrsg.), *Basiswissen Lehrerbildung* (S. 207–223). Klett.
- Kleickmann, T., Steffensky, M. & Praetorius, A.-K. (2020). Quality of Teaching in Science Education. More Than Three Basic Dimensions? *Zeitschrift für Pädagogik*, 66(Beiheft 1/20), 37–55.
- Klein, J. (2009). Erklären-Was, Erklären-Wie, Erklären-Warum. Typologie und Komplexität zentraler Akte der Welterschließung. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (2. Aufl., S. 25–36). Tübingen: Stauffenberg.
- Klieme, E. (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einleitung in den Thementeil. *Zeitschrift für Pädagogik*, 52(6), 765–773.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). Zur Entwicklung nationaler Bildungsstandards. BMBF (Bildungsforschung, 1). http://sinus-transfer.uni-bayreuth.de/fileadmin/MaterialienBT/Expertise_Bildungsstandards.pdf [07.08.2022].
- Klieme, E., Lipowsky, F. & Rakoczy, K. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Waxmann.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik: Outcomeorientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54(2), 222–237.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I. „Aufgabenkultur“ und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 43–57). BMBF.
- KMK (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung: Beschluss der KMK vom 16.10.2008 i. d. F. vom 14.3.2019*, Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK). https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf [07.08.2022].
- Knigge, J. (2011). *Modellbasierte Entwicklung und Analyse von Testaufgaben zur Erfassung der Kompetenz „Musik wahrnehmen und kontextualisieren“*. LIT
- Knigge, J. (2014). Der Kompetenzbegriff in der Musikpädagogik: Verwendung, Kritik, Perspektive. In J. Vogt, F. Heß & M. Brenk (Hrsg.), *(Grund-)Begriffe musikpädagogischen Nachdenkens. Entstehung, Bedeutung, Gebrauch* (S. 105–135). LIT.
- Knigge, J., Lehmann, A. C., Lehmann-Wermser, A. & Niessen, A. (2008). Bildungsstandards und Kompetenzmodelle für das Fach Musik? *Zeitschrift für Kritische Musikpädagogik, Sonderedition 2*, 1–2.

- Knutson, K. & Crowley, K. (2010). Connecting with Art: How Families Talk About Art in a Museum Setting. In M. K. Stein & L. Kucan (Hrsg.), *Instructional Explanations in the Disciplines* (S. 190–206). Springer. https://doi.org/10.1007/978-1-4419-0594-9_12
- Kohl, K. (2007). *Metapher*. Springer. <https://doi.org/10.1007/978-3-476-05075-5>
- Kotthoff, H. (2009). Gesprächsfähigkeiten: Erzählen, Argumentieren, Erklären. In M. Krelle & C. Spiegel (Hrsg.), *Sprechen und Kommunizieren* (S. 41–63). Schneider Verlag Hohengehren.
- Kraemer, R.-D. (2007). *Musikpädagogik – eine Einführung in das Studium* (2., verb. Aufl.). Wißner.
- Krämer, O. (2007). Musik malen. In N. Heukäufer (Hrsg.), *MusikMethodik. Handbuch für die Sekundarstufe I und II* (S. 240–248). Cornelsen Scriptor.
- Krämer, O. (2018a). Improvisation als didaktisches Handlungsfeld. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 319–326). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Krämer, O. (2018b). Musik und andere künstlerische Ausdrucksformen als didaktisches Handlungsfeld. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 341–348). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Kranefeld, U. (2008). Zwischen explorativem Musizieren und ästhetischer Reflexion. Ergebnisse einer Studie zu Gruppenkompositionsprozessen zu Bildern im Musikunterricht der gymnasialen Oberstufe. In A. C. Lehmann & M. Weber (Hrsg.), *Musizieren innerhalb und außerhalb der Schule* (S. 159–173). Die Blaue Eule.
- Kranefeld, U. (2015). Ergebnisse der Forschungen zu den Programmen *Jedem Kind ein Instrument* in Nordrhein-Westfalen und Hamburg. In U. Kranefeld (Hrsg.), *Instrumentalunterricht in der Grundschule. Prozess- und Wirkungsanalysen zum Programm Jedem Kind ein Instrument* (S. 6–18). BMBF.
- Kranefeld, U. (2021). Der Diskurs um Unterrichtsqualität in der Musikdidaktik zwischen generischen und fachspezifischen Dimensionen. *Unterrichtswissenschaft*, 49(2), 221–233. <https://doi.org/10.1007/s42010-021-00113-y>
- Kranefeld, U., Busch, T. & Dücker, J. (2015). BEGIn – Instrumentaler Gruppenunterricht in der Grundschule: Teilnahme, Selbstkonzepte, Gestaltungsmuster. In U. Kranefeld (Hrsg.), *Instrumentalunterricht in der Grundschule. Prozess- und Wirkungsanalysen zum Programm Jedem Kind ein Instrument* (S. 49–89). BMBF.
- Kranefeld, U., Heberle, K. & Naacke, S. (2015). Videografische Befunde zu Aspekten von Unterrichtsqualität im ersten JeKi-Jahr. In U. Kranefeld (Hrsg.), *Instrumentalunterricht in der Grundschule. Prozess- und Wirkungsanalysen zum Programm Jedem Kind ein Instrument* (S. 148–166). BMBF.
- Kranefeld, U. Mause, A.-L. & Duve, J. (2019). Zur Materialität von Prozessen des Musik-Erfindens: Interaktionsanalytische Zugänge zur Wandelbarkeit der Dinge. In V. Weidner (Hrsg.), *Praxen und Diskurse aus Sicht musikpädagogischer Forschung*. Waxmann. <https://doi.org/10.25656/01:20703>
- Kranich, W. (2016). *Sprechwissenschaftliche Grundlagen der Prosodieperzeption*. Frank & Timme.
- Krause-Benz, M. (2016). Handlungsorientierung – zwischen Praktizismus und Performativität. In Lehmann-Wermser, A. (Hrsg.), *Musikdidaktische Konzeptionen. Ein Studienbuch* (S. 83–96). Wißner.
- Krauss, S. (2020). Expertise-Paradigma in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 154–162). Klinkhardt/UTB. <https://www.utb.de/doi/book/10.36198/9783838554730>

- Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., Besser, M. & Elsner, J. (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). Waxmann.
- Krauss, S. & Bruckmaier, G. (2014). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. und erw. Aufl., S. 241–261). Waxmann.
- Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N. & Blum, W. (2020). Competence as a continuum in the COACTIV study: the “cascade model”. *ZDM Mathematics Education*, 52(2), 311–327. <https://doi.org/10.1007/s11858-020-01151-z>
- Krauss, S., Lindl, A., Schilcher, A., Fricke, M., Göhring, A., Hofmann, B., Kirchhoff, P., Mulder, R. H. (Hrsg.) (2017). *FALKO: Fachspezifische Lehrerkompetenzen: Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik*. Waxmann.
- Krauss, S. & Puffer, G. (2019). „Wurzeln“ und „Brüche“ in der Entwicklung eines Tonsystems. Musik und Mathematik. In G. Puffer, A. Becker, F. Körndle & K. Sprau (Hrsg.), *Musik – Pädagogik – Professionalität. Festschrift für Bernhard Hofmann zum 60. Geburtstag* (S. 303–319). Helbling.
- Krupp, V. (2021). *Wirksamer Musikunterricht*. Schneider Verlag Hohengehren.
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (3. Aufl.). Beltz Juventa.
- Kühn, C. (2002). Musiktheorie zwischen Hochschule und Schule. *Diskussion Musikpädagogik*, 13, 6–21.
- Kühn, C. (2006). *Musiktheorie unterrichten – Musik vermitteln*. Bärenreiter.
- Kulgemeyer, C. (2019). Towards a framework for effective instructional explanations in science teaching. *Studies in Science Education*, 54(2), 109–139. <https://doi.org/10.1080/03057267.2018.1598054>
- Kulgemeyer, C. & Schecker, H. (2013). Students Explaining Science – Assessment of Science Communication Competence. *Research in Science Education*, 43(6), 2235–2256. <https://doi.org/10.1007/s11165-013-9354-1>
- Kulgemeyer, C. & Tomczyszyn, E. (2015). Physik erklären – Messung der Erklärens-fähigkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften*, 21, 111–126. <https://doi.org/10.1007/s40573-015-0029-5>
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environment Research*, 9, 231–251. <https://doi.org/10.1007/s10984-006-9015-7>
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Waxmann. <https://doi.org/10.31244/9783830974338>
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013). Professional competence of teachers: effects on instructional quality and student development. *Journal of educational psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts*. Brill, Schöningh/UTB. <https://doi.org/10.36198/9783838538952>
- Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85–113). Waxmann. <https://doi.org/10.31244/9783830974338>

- Lang, R. (2008). Musiktheorie im Unterricht. Impulse für eine umstrittene Fachdisziplin. In F. Heß, & J. Terhag (Hrsg.), *Musikunterricht heute 7. Bach – Bebop – Bredemeyer. Sperriges lebendig unterrichten* (S. 212–222). Lugert Verlag.
- Lang, R. (2010). Musiktheorie in musizierpraktischem Schulunterricht. Zur Effizienz basaler Harmonielehre für das Improvisieren mit Keyboards. In N. Knolle (Hrsg.), *Evaluationsforschung in der Musikpädagogik* (S. 255–273). Die Blaue Eule.
- Lauer, D. (2018). Musikbezogenes Wahrnehmen als didaktisches Handlungsfeld. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 311–315). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Lawrence, M. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. [Software] <https://github.com/mike-lawrence/ez> (R package version 4.4-0).
- Lazaraton, A. (2004). Gesture and Speech in the Vocabulary Explanations of One ESL Teacher: A Microanalytic Inquiry. *Language Learning*, 54(1), 79–117. <https://doi.org/10.1111/j.1467-9922.2004.00249.x>
- Leeb, E.-M. (2021). *JugendChorLeitung. Impulse für eine ansprechende und erfolgreiche Jugendchorarbeit*. Con Brio.
- Lehmann-Wermser, A. & Krause-Benz, M. (2013). Musiklehrer(-Bildung) im Fokus musikpädagogischer Forschung – ein Vorwort. In A. Lehmann-Wermser & M. Krause-Benz (Hrsg.), *Musiklehrer(-bildung) im Fokus musikpädagogischer Forschung* (S. 7–12). Waxmann.
- Lehmann-Wermser, A., Weishaupt, H. & Konrad, U. (2020). *Musikunterricht in der Grundschule. Aktuelle Situation und Perspektive*. Bertelsmann Stiftung.
- Lehner, M. (2018). *Erklären und Verstehen: Eine kleine Didaktik der Vermittlung*. Haupt/UTB. <https://doi.org/10.36198/9783838548920>
- Lehner, M. (2020). *Didaktische Reduktion* (2. Aufl.). Haupt/UTB. <https://doi.org/10.36198/9783838553832>
- Leinhardt, G. (2001). Instructional Explanations: A Commonplace for Teaching and Location for Contrast. In V. Richardson (Hrsg.), *Handbook of research on teaching* (S. 333–357). American Educational Research Association.
- Leisen, J. (2007). Das Erklären im Unterricht. *Der Mathematische und Naturwissenschaftliche Unterricht*, 60(8), 459–462.
- Leiss, D., Domenech, M., Ehmke, T. & Schwippert, K. (2017). Schwer – schwierig – diffizil: Zum Einfluss sprachlicher Komplexität von Aufgaben auf fachliche Leistungen in der Sekundarstufe 1. In D. Leiss, M. Hagen, A. Neumann & K. Schwippert (Hrsg.), *Sprache im Fach Mathematik – Forschungsstand und Herausforderungen im Verlauf der Schulzeit* (S. 99–126). Waxmann.
- Lenke, S. (2006). Die Funktionskreise Respiration, Phonation, Artikulation – Auffälligkeiten bei Lehramtsstudierenden. *Sprache – Stimme – Gehör*, 30, 24–28. <https://doi.org/10.1055/s-2006-931529>
- Lenord, C. (2020). Professionelle Wahrnehmung von Musikunterricht durch Unterrichtsvideos – kreativ und strukturiert. In K. Kaspar, M. Becker-Mrotzek, S. Hofhues, J. König & D. Schmeinck (Hrsg.), *Bildung. Schule, Digitalisierung* (S. 247–252). Waxmann.
- Lenske, G. (2016). *Schülerfeedback in der Grundschule – Untersuchung zur Validität*. Waxmann.
- Leonhart, R. (2015). *Lehrbuch Statistik* (4. überarb. und erw. Aufl.). Hogrefe.
- Lessing, W. & Stöger, C. (2018). Lehrende in musikpädagogischen Arbeitsfeldern. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 132–142). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>

- Leuders, T. (2015). Aufgaben in Forschung und Praxis. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 435–460). Springer. https://doi.org/10.1007/978-3-642-35119-8_16
- Lill, F., Hasselhorn, J. & Lehmann, A. C. (2019). Der Zusammenhang von musikalischem Selbstkonzept und musikpraktischen Kompetenzen in der Sekundarstufe I. In V. Weidner & C. Rolle (Hrsg.), *Praxen und Diskurse aus Sicht musikpädagogischer Forschung* (S. 171–187). Waxmann.
- Lipowsky, F. (2007). Was wissen wir über guten Unterricht? Im Fokus: die fachliche Lernentwicklung. In G. Becker, A. Feindt, H. Meyer, M. Rothland, L. Stäudel & E. Terhart (Hrsg.), *Guter Unterricht. Maßstäbe & Merkmale – Wege und Werkzeuge*. Friedrich Jahresheft XXV (S. 26–30). Friedrich Verlag.
- Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 73–102). Springer. https://doi.org/10.1007/978-3-540-88573-3_4
- Lindl, A., Gaier, L., Weich, M., Frei, M., Ehras, C., Gastl-Pischetsrieder, M., Elmer, M., Asenmolz, K., Ruck, A.-M. & Heinze, J. (2019). Eine ‚gute‘ Erklärung für alle?! Gruppenspezifische Unterschiede in der Beurteilung von Erklärqualität – Erste Ergebnisse aus dem interdisziplinären Forschungsprojekt FALKE. In M. Pietsch, T. Ehmke & P. Kuhl (Hrsg.), *Lehrer. Bildung. Gestalten*. (S. 128–141). Beltz.
- Lindl, A. & Hilbert, S. (i. Vorb.). Transdisziplinäre Perspektiven auf Erklären im Unterricht. In A. Schilcher, A. Lindl, S. Krauss & S. Hilbert (Hrsg.), *Fachspezifische Lehrerkompetenzen im Erklären*. Beltz Juventa.
- Lindl, A. & Kloiber, H. (2017). FALKO-L: Modellierung und Messung domänenspezifischer Kompetenzen von Lateinlehrkräften. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 153–199). Waxmann.
- Lindl, A. & Krauss, S. (2017). Transdisziplinäre Perspektiven auf domänenspezifische Lehrerkompetenzen. Eine Metaanalyse zentraler Resultate des Forschungsprojektes FALKO. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 381–438). Waxmann.
- Linn, F. (2017). *Überzeugungen von Musiklehrenden zum Umgang mit Heterogenität im Musikunterricht*. universi – Universitätsverlag Siegen.
- Lotz, M. & Lipowsky, F. (2015). Die Hattie-Studie und ihre Bedeutung für den Unterricht – Ein Blick auf ausgewählte Aspekte der Lehrer-Schüler-Interaktion. In G. Mehlhorn, F. Schulz & K. Schöppe (Hrsg.), *Begabungen entwickeln & Kreativität fördern* (S. 97–136). Kopaed.
- Lukaschyk, J. (2018). Prävention von Stimmstörungen bei Lehrkräften. Stimmseminare und Coaching – Erfahrungen und Empfehlungen. *Forum Logopädie*, 32(6), 6–10.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. Aufl.). Beltz.
- Mayring, P. (2016). *Einführung in die qualitative Sozialforschung: eine Anleitung zu qualitativem Denken* (6. Aufl.). Beltz.
- Meyer, H. (2004). *Was ist guter Unterricht?* Cornelsen.
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2019). *Kernlehrplan für die Sekundarstufe I Gymnasium in Nordrhein-Westfalen Musik*. Düsseldorf. https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/207/g9_mu_klp_%203406_2019_06_23.pdf [07.08.2022].
- Mohr, A. (2014). *Handbuch der Kinderstimmgebung*. Schott.

- Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K. & Ly, A. (2022). BayesFactor: Computation of Bayes Factors for Common Designs. [Software] <https://richarddmorey.github.io/BayesFactor/> (R package version 0.9.12-4.4).
- Mulder, R. & 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* (S. 427–438). Springer. https://doi.org/10.1007/978-3-531-94025-0_30
- Mulder, R., Sauer, S. & Kempka, F. (2017). FALKO-PA: Ein Instrument aus flexibel einsetzbaren Vignetten zur Erfassung pädagogischer Kompetenzen. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 337–379). Waxmann.
- Nathan, M. J. & Petrosino, A. (2003). Expert blind spot among pre-service teachers. *American Educational Research Journal*, 40, 905–928. <https://doi.org/10.3102/00028312040004905>
- Neubrand, M., Jordan, A., Krauss, S., Blum, W. & Löwen, K. (2011). Aufgaben im COACTIV-Projekt: Einblicke in das Potenzial für kognitive Aktivierung im Mathematikunterricht. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 115–132). Waxmann.
- Neuhäus, D. (2008). *Perspektive Musiklehrer/in. Der Berufswahlprozess von Lehramtsstudierenden mit dem Unterrichtsfach Musik*. Verlag Dohr.
- Neuhäus, D. (2019). Konzepte zur Förderung von Reflexionsfähigkeit als Beitrag zur Professionalisierung aus musikpädagogischer Sicht. In G. Puffer, A. Becker, F. Körndle & K. Sprau (Hrsg.), *Musik – Pädagogik – Professionalität. Festschrift für Bernhard Hofmann zum 60. Geburtstag* (S. 207–219). Helbling.
- Neumeister, N. (2011). *(Wie) Wird im Deutschunterricht erklärt? – Wissensvermittelnde Handlungen im Sprachunterricht der Sekundarstufe I* (Dissertation). https://phbl-opus.phlb.de/frontdoor/deliver/index/docId/32/file/1_Dissertation.pdf [07.08.2022].
- Neumeister, N. & Vogt, R. (2009). Erklären im Unterricht. In M. Becker-Mrotzek (Hrsg.), *Handbuch Mündlichkeit* (S. 540–561). Schneider Verlag Hohengehren.
- Neuweg, G. H. (2014). Das Wissen der Wissensvermittler. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. und erw. Aufl., S. 583–614). Waxmann.
- Niessen, A. (2006a). Individualkonzepte von MusiklehrerInnen. Ein qualitativer Ansatz in der musikpädagogischen Lehr-/Lernforschung. In N. Knolle (Hrsg.), *Lehr- und Lernforschung in der Musikpädagogik* (S. 175–199). Blaue Eule.
- Niessen, A. (2006b). *Individualkonzepte von Musiklehrern*. LIT.
- Niessen, A. (2007). Musikalische und pädagogische Selbstkonzepte von Musiklehrerinnen und Musiklehrern. *Diskussion Musikpädagogik*, 33, 30–40.
- Niessen, A. (2010). Die Bedeutung von Verarbeitungstiefe im Musikunterricht – Dimensionen von Unterrichtsqualität einer Musikstunde. In C. Wallbaum (Hrsg.), *Perspektiven der Musikdidaktik. Drei Schulstunden im Licht der Theorien* (S. 63–82). Olms.
- Niessen, A. (2013). Das Unterrichten großer Lerngruppen im ersten JeKi-Jahr aus Lehrendenperspektive. In A. Lehmann-Wermser & M. Krause-Benz (Hrsg.), *Musiklehrer(-bildung) im Fokus musikpädagogischer Forschung* (S. 81–97). Waxmann.
- Niessen, A. & Lehmann-Wermser, A. (2006). Musikunterricht im Spiegel mehrperspektivischer Sinnzuschreibungen. In N. Knolle (Hrsg.), *Lehr- und Lernforschung in der Musikpädagogik* (S. 239–252). Die Blaue Eule.

- Niessen, A., Lehmann-Wermser, A., Knigge, J. & Lehmann, A. C. (2008). Entwurf eines Kompetenzmodells ‚Musik wahrnehmen und kontextualisieren‘. *Zeitschrift für kritische Musikpädagogik, Sonderedition 2*, 3–33.
- Nimczik, O. (1997). Musik erfinden. In S. Helms, R. Schneider & R. Weber (Hrsg.), *Handbuch des Musikunterrichts. Sekundarstufe II* (S. 169–188). Bosse.
- Nimczik, O. (2002). Gedanken zum Singen in der Schule. *Musik und Bildung*, 34(3), 4–6.
- Nimczik, O. (2013). Neue Musik in der Schule. In W. Jank (Hrsg.), *MusikDidaktik. Praxishandbuch für die Sekundarstufe I und II* (5. Aufl., S. 195–203). Cornelsen Scriptor.
- Nowak, C. (1999). *Elementare Musiklehre und Grundlagen der Harmonielehre. Begleitbuch für den modernen Musikunterricht, mit praktischen Übungsbeispielen*. Edition DUX.
- Nückles, M., Wittwer, J. & Renkl, A. (2005). Information about a layperson’s knowledge supports experts in giving effective and efficient online advice to laypersons. *Journal of Experimental Psychology: Applied*, 11, 219–236. <https://doi.org/10.1037/1076-898X.11.4.219>
- Nusseck, M., Immerz, A., Spahn, C., Echternach, M. & Richter, B. (2019). Long-Term Effects of a Voice Training Programm for Teachers on Vocal and Mental Health. *Journal of Voice*, 35(3), 438–446. <https://doi.org/10.1016/j.voice.2019.11.016>
- Oberhaus, L. (2007). Neues vom Musikpädagogischen Eros. (Un)zeitgemäße Betrachtungen zur ‚Musiklehrerpersönlichkeit‘ anhand verschiedener Musiklehrerrollen im Film. *Zeitschrift für Kritische Musikpädagogik 2007*, 72–85. <https://www.zfkm.org/07-oberhaus.pdf> [07.08.2022].
- Oberhaus, L. (2015). Über Musik reden. Darstellung und Vergleich von fünf Ansätzen musikbezogenen Erzählens im Unterricht. *Zeitschrift für Kritische Musikpädagogik, Sonderedition 3*, 49–70.
- Oberschmidt, J. (2010). Metaphorischer Sprachgebrauch im Unterricht. Überlegungen zur Evaluierung der Schülersprache. In N. Knolle (Hrsg.), *Evaluationsforschung in der Musikpädagogik* (S. 131–154). Die Blaue Eule.
- Odora, R. J. (2014). Using Explanation as a Teaching Method: How prepared are High School Technology Teachers in Fredd State Province, South Africa? *Journal of Social Science*, 38(1), 71–81. <https://doi.org/10.1080/09718923.2014.11893238>
- Oebelsberger, M. (2019). Elementarisierung. Notwendigkeiten und Grenzen eines Unterrichtsprinzips für den Musikunterricht. In A. M. Kalcher & M. Oebelsberger (Hrsg.), *Elementar. Künstlerisch-Pädagogische Sichtweisen* (S. 86–94). LIT.
- Overmann, U. (2002). Professionalisierungsbedürftigkeit und Professionalisiertheit pädagogischen Handelns. In M. Kraul (Hrsg.), *Biographie und Profession* (S. 19–63). Klinkhardt.
- Ollesch, J., Dörfler, T. & Vogel, M. (2018). Die inhaltliche Validierung von Unterrichtsvignetten durch eine mehrstufige Expertenbefragung. In J. Rutsch, M. Rehm, M. Vogeck, M. Seidenfuß & T. Dörfler (Hrsg.), *Effektive Kompetenzdiagnose in der Lehrerbildung. Professionalisierungsprozesse angehender Lehrkräfte untersuchen* (S. 129–151). Springer. https://doi.org/10.1007/978-3-658-20121-0_7
- Pabst-Krueger, M. (2013). Klassenmusizieren. In W. Jank (Hrsg.), *MusikDidaktik. Praxishandbuch für die Sekundarstufe I und II* (5., überarb. Neuaufl., S. 158–168). Cornelsen Scriptor.
- Pauli, C. (2015). Einen Sachverhalt erklären. *Pädagogik*, 3, 44–47.
- Pendorf, T. (2019). „Es ist ein sehr lautes Lied und am Anfang kommt es in Pausen“. Zur Relevanz eines sprachsensiblen Musikunterrichts. *VDS Magazin*, 40(2), 8–13.
- Peterson, D. (1996). *Forms of Representation*. Intellect.
- Petrasch, T. & Zinke, J. (2012). *Videofilm. Konzeption und Produktion* (2. Aufl.). Hanser. <https://doi.org/10.3139/9783446431584>
- Pezenburg, M. (2018). Singen als didaktisches Handlungsfeld. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 294–299). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>

- Pfeiffer, W. (1994). *Musiklehrer: Biographie, Alltag und berufliche Zufriedenheit von Musiklehrern an bayerischen Gymnasien. Eine theoretische und empirische Analyse*. Die Blaue Eule.
- Pianta, R. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38, 109–119. <https://doi.org/10.3102/0013189X09332374>
- Pissarek, M. & Schilcher, A. (2017). FALKO-D: Die Untersuchung des Professionswissens von Deutschlehrenden. Entwicklung eines Messinstruments zur fachspezifischen Lehrerkompetenz und Ergebnisse zu dessen Validierung. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 67–111). Waxmann.
- Piowar, V., Barth, V. L., Ophardt, D. & Thiel, V. (2018). Evidence-Based Scripted Videos on Handling Student Misbehavior: The Development and Evaluation of Video Cases for Teacher Education. *Professional Development in Education*, 44(3), 369–384. <https://doi.org/10.1080/19415257.2017.1316299>
- Platz, F. & Lehmann, A. C. (2018). Üben aus musikpsychologischer Perspektive. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 268–275). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Praetorius, A.-K. (2013). Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter. Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31(2), 174–185. <https://doi.org/10.36950/bzl.31.2013.9650>
- Praetorius, A.-K. & Charalambous, C.Y. (2018). Classroom observation frameworks for studying teaching quality: looking back and looking forward. *ZDM Mathematics Education*, 50(3), 535–553. <https://doi.org/10.1007/s11858-018-0946-0>
- Praetorius, A.-K. & Gräsel, C. (2021). Noch immer auf der Suche nach dem heiligen Gral: Wie generisch oder fachspezifisch sind Dimensionen der Unterrichtsqualität? *Unterrichtswissenschaft*, 49, 167–188. <https://doi.org/10.1007/s42010-021-00119-6>
- Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülzdorf-Kersting, M., Heinitz, B. & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft*, 48, 409–446. <https://doi.org/10.1007/s42010-020-00082-8>
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of three basic dimensions. *ZDM Mathematics Education*, 50(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Praetorius, A.-K. & Nehring, A. (2020). Unterrichtsqualität zwischen Generik und Fachspezifik: Einführung in den Thementeil. *Unterrichtswissenschaft*, 48, 297–301. <https://doi.org/10.1007/s42010-020-00081-9>
- Praetorius, A.-K., Rogh, W. & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft*, 48, 303–318. <https://doi.org/10.1007/s42010-020-00072-w>
- Prantl, D. & Wallbaum, C. (2017). Der Analytical Short Film in der Lehrerbildung. Darstellung einer Seminarmethode und Kurzbericht einer wissenschaftlichen Begleitforschung an der Hochschule für Musik und Theater Leipzig. In A. J. Cvetko & C. Rolle (Hrsg.), *Musikpädagogik und Kulturwissenschaft* (S. 289–308). Waxmann.

- Prediger, S. & Erath, K. (2014). Content, Interaction, or Both? Synthesizing Two German Traditions in a Video Study on Learning to Explain in Mathematics Classroom Microcultures. *EUR-ASIA Journal of Mathematics, Science & Technology Education*, 10(4), 313–327. <https://doi.org/10.12973/eurasia.2014.1085a>
- Prinz, U. & Scheyett, A. (1993). *Musik um uns 2. Allgemeine Ausgabe für das 7. – 10. Schuljahr* (3. Aufl.). Metzler.
- Prüfer, P. & Rexroth, M. (2005). *Kognitive Interviews*. Mannheim: Zentrum für Umfragen, Methoden und Analysen (ZUMA). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-201470> [07.08.2022].
- Puffer, G. (2013). Musikunterricht im Rückspiegel. Zur Konstruktion von „Musikunterricht“ in Stundennachbesprechungen während der ersten Ausbildungsphase. In A. Lehmann-Wermser & M. Krause-Benz (Hrsg.), *Musiklehrer(-bildung) im Fokus musikpädagogischer Forschung* (S. 29–44). Waxmann.
- Puffer, G. (2019). „Lieder nicht zu Hörbeispielen degradieren“: Der Unterrichtsgegenstand „Musik“ in Stundennachbesprechungen eines Theorie-Praxis-Seminars. In G. Puffer, A. Becker, F. Körndle & K. Sprau (Hrsg.), *Musik – Pädagogik – Professionalität. Festschrift für Bernhard Hofmann zum 60. Geburtstag* (S. 221–243). Helbling.
- Puffer, G. (2021). Professionelle Kompetenzen von Musiklehrkräften: Grundzüge einer Modellierung. *Beiträge Empirischer Musikpädagogik*, 12, 1–71. <https://www.b-em.info/index.php/ojs/article/view/207> [07.08.2022].
- Puffer, G. (2022). Singen im schulischen Musikunterricht: Professionswissen als Basis von Handlungsentscheidungen. In M. Göllner, J. Knigge A. Niessen & V. Weidner (Hrsg.), *43. Jahressband des Arbeitskreises Musikpädagogische Forschung [43rd Yearbook of the German Association for Research in Music Education]*. Waxmann.
- Puffer, G. & Hofmann, B. (2016). FALKO-M: Zur Konzeptualisierung des Professionswissens von Musiklehrkräften. In J. Knigge & A. Niessen (Hrsg.), *Musikpädagogik und Erziehungswissenschaft* (S. 107–120). Waxmann.
- Puffer, G. & Hofmann, B. (2017). FALKO-M: Entwicklung und Validierung eines Testinstruments zum domänenspezifischen Professionswissen von Musiklehrkräften. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 245–289). Waxmann.
- Puffer, G. & Hofmann, B. (2022). Professionelle Kompetenz(en) von Musiklehrkräften: Ein empirisch begründetes Modell. *Zeitschrift für Erziehungswissenschaften* 2022. <https://doi.org/10.1007/s11618-022-01101-3>
- R Core Team (2021). base: The R Base Package. [Software] <https://www.r-project.org> (R package version 4.1.2).
- Rathausky, A. (2010). *Erklärprozesse im Fach Englisch. Eine qualitative Studie zur Vermittlung grammatikalischer Inhalte in der Sekundarstufe*. [Dissertation]. <https://phbl-opus.phlb.de/frontdoor/deliver/index/docId/31/file/DoktorarbeitRathausky.pdf> [07.08.2022].
- Rauhe, H., Reinecke, H.-P. & Ribke, W. (1975). *Hören und Verstehen. Theorie und Praxis handlungsorientierten Musikunterrichts*. Kösel-Verlag.
- Rehbein, J. (1977). *Komplexes Handeln Elemente zur Handlungstheorie der Sprache*. Metzler. <https://doi.org/10.1007/978-3-476-03078-8>
- Rehm, M. & Bölsterli, K. (2014). Entwicklung von Unterrichtsvignetten. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 213–225). Springer Spektrum. https://doi.org/10.1007/978-3-642-37827-0_18

- Reid, D. A. & Roberts, R. (2004). Adult learners' criteria für explanations. *ZDM Mathematics Education*, 36, 140–149. <https://doi.org/10.1007/BF02655666>
- Reinecke, H.-P. (1974). Die Sprachebene über Musik als Hierarchie relationaler Systeme. Überlegungen zu Wittgensteins ‚Sprachstil-Modell‘. In H. H. Eggebrecht (Hrsg.), *Zur Terminologie der Musik des 20. Jahrhunderts* (S. 22–32). Musikwissenschaftliche Verlagsgesellschaft.
- Reiss, K., Weis, M., Klieme, E. & Köller, O. (Hrsg.) (2019). *PISA 2018. Grundbildung im internationalen Vergleich*. Waxmann. <https://doi.org/10.31244/9783830991007>
- Renkl, A. (2017). Instruktionale Erklärungen. In M. A. Wirtz (Hrsg.), *Lexikon der Psychologie* (S. 799–800). Hogrefe.
- Renkl, A., Wittwer, J., Große, C., Hauser, S., Hilbert, T., Nückles, M. & Schworm, S. (2006). Instruktionale Erklärungen beim Erwerb kognitiver Fertigkeiten: sechs Thesen zu einer oft vergeblichen Bemühung. In I. Hosenfeld & F.W. Schrader (Hrsg.), *Schulische Leistung. Grundlagen, Bedingungen, Perspektiven* (S. 205–223). Waxmann.
- Reusser, K. (2008). Empirisch fundierte Didaktik – didaktisch fundierte Unterrichtsforschung. *Zeitschrift für Erziehungswissenschaft, Sonderheft 9*, 219–237. https://doi.org/10.1007/978-3-531-91775-7_15
- Reusser, K. & Pauli, C. (2003). *Mathematikunterricht in der Schweiz und in weiteren sechs Ländern. Bericht über die Ergebnisse einer internationalen und schweizerischen Video-Unterrichtsstudie*. Universität Zürich: Pädagogisches Institut.
- Richter, C. (1976). *Theorie und Praxis der didaktischen Interpretation von Musik*. Verlag Moritz Diesterweg.
- Richter, C. (1993). Lebensweltliche Orientierung des Musikunterrichts. *Musik und Bildung*, 6, 24–29.
- Richter, C. (2002). Musiktheorie im Musiklehrerstudium und im Musikunterricht. *Diskussion Musikpädagogik*, 13, 45–55.
- Riese, J., Kulgemeyer, C., Zander, S., Borowski, A., Fischer, H. E., Gramzow, Y., Reinhold, P., Schecker, H. & Tomczyszyn, E. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden* (S. 55–79). Beltz Juventa.
- Röbke, P. (2018). Dimensionen des künstlerischen aus instrumentalpädagogischer Perspektive. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 192–197). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Rogerson, J. & Dodd, B. (2005). Is There an Effect of Dysphonic Teacher's Voices on Children's Processing of Spoken Language? *Journal of Voice*, 19(1), 47–60. <https://doi.org/10.1016/j.voice.2004.02.007>
- Rolle, C. (1999) *Musikalisch-ästhetische Bildung: Über die Bedeutung ästhetischer Erfahrung für musikalische Bildungsprozesse*. Bosse.
- Rolle, C. (2008) Musikalische Bildung durch Kompetenzerwerb? Überlegungen im Anschluss an den Entwurf eines Kompetenzmodells „Musik wahrnehmen und kontextualisieren“. *Zeitschrift für Kritische Musikpädagogik*, 2008, 42–59. <http://www.zfkm.org/sonder08-rolle.pdf> [07.08.2022].
- Rolle, C. (2010). Musikdidaktische Reflexionen: Was heißt musikalische Bildung durch Inszenierung ästhetischer Erfahrungsräume? In C. Wallbaum (Hrsg.), *Perspektiven der Musikdidaktik* (S. 197–223). Olms.
- Rolle, C. (2011). Sprecht miteinander! Reden über Musik ist nicht das Andere des Klassenmusizierens. *Diskussion Musikpädagogik, Sonderheft 3*, 130–133.

- Rolle, C. (2014). Ästhetischer Streit als Medium des Musikunterrichts – zur Bedeutung des argumentierenden Sprechens über Musik für ästhetische Bildung. *Art Education Research*, 5(9). https://blog.zhdk.ch/iaejournal/2014/12/16/n9_aesthetischer-streit-als-medium-des-musikunterrichts-zur-bedeutung-des-argumentierenden-sprechens-ueber-musik-fuer-aesthetische-bildung/ [07.08.2022].
- Rösler, F. (2011). *Psychologie der Kognition. Eine Einführung in die Kognitive Neurowissenschaft*. Springer Spektrum. <https://doi.org/10.1007/978-3-8274-2599-7>
- Rothland, M. (2014). Warum entscheiden sich Studierende für den Lehrerberuf? In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. und erw. Aufl., S. 349–385). Waxmann.
- Rucha, B. (2016). *Crashkurs Dirigieren. Schlagtechnik – Probenarbeit – Repertoire – Führungspersönlichkeit*. Schott.
- Ruf, B. (2014). Wie denken Lehrer*innen über (das Unterrichten von) Musiktheorie? In B. Clausen (Hrsg.), *Teilhabe und Gerechtigkeit* (S. 131–143). Waxmann.
- Schäfer, C. (2021). *Dirigieren Plus. Dirigiertechnik – Persönlichkeitsentwicklung – (Selbst-)Organisation*. Bärenreiter.
- Schäfer-Lembeck, H.-U. (2003). Begriff und Praxis. Einblicke in eine Fortbildungsveranstaltung. In H. Baumann (Hrsg.), *Zwischen Nützlichkeitsdenken und kulturellem Auftrag. Musikunterricht für die Schule des 21. Jahrhunderts* (S. 205–214). Kongressbericht „Tage der Bayerischen Schulmusik 2003“ [Verband bayerischer Schulmusiker e. V.].
- Scheffel, L. (2019). *Erklären im Unterricht. Taktiken und Strategien*. Beltz.
- Schilcher, A. & Krauss, S. (2019). Anforderungen an die Lehrerbildung aus fachdidaktischer Perspektive. In N. McElvany, W. Bos, H. G. Holtappels & A. Ohle-Peters (Hrsg.), *Bedingungen und Effekte von Lehrerbildung, Lehrkraftkompetenzen und Lehrkraft Handeln* (S. 29–49). Waxmann.
- Schilcher, A., Krauss, S., Kirchhoff, P., Lindl, A., Hilbert, S., Asen-Molz, K., Ehras, C., Elmer, M., Frei, M., Gaier, L., Gastl-Pischetsrieder, M., Gunga, E., Murmann, R., Röhr, S., Ruck, A.-M., Weich, M., Dittmer, A., Fricke, M., Hofmann, B., Memminger, J., Rank, A., Tepner, O. & Thim-Mabrey, C. (2021). FALKE: Experiences from Transdisciplinary Educational Research by Fourteen Disciplines. *Frontiers in Education*, 5. <https://doi.org/10.3389/educ.2020.579982>
- Schilcher, A., Krauss, S., Rincke, K. & Hilbert, S. (2017). Ausblick – Aus FALKE wird FALKE. Fachspezifische Lehrerkompetenz im Erklären. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKE: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 439–451). Waxmann.
- Schilcher, A., Lindl, A., Krauss, S. & Hilbert, S. (i. Vorb.) *Fachspezifische Lehrerkompetenzen im Erklären*. Beltz Juventa.
- Schlothfeldt, M. (2018). Komposition als didaktisches Handlungsfeld. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 326–333). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Schmid, W., Hofmann, B., Lindner, U., Mack, D., Mastnak, W. & Zink, B. (2019). *Tonart 7/8. Musik erleben – reflektieren – interpretieren. Lehrband (Regionalausgabe Bayern)*. Helbling.
- Schmidt, A. (2016). *Ästhetische Erfahrung in Gesprächen über Kunst. Eine empirische Studie mit Fünft- und Sechstklässlern*. Kopaed.
- Schmölzer-Eibinger, S. (2013). Sprache als Medium des Lernens im Fach. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach: Sprachlichkeit und fachliches Lernen* (S. 25–40). Waxmann.
- Schnell, R., Hill, P. B. & Esser, E. (2018). *Methoden der empirischen Sozialforschung*. De Gruyter.

- Schnotz, W. & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen. *Experimental Psychology*, 46(3), 217–236. <https://doi.org/10.1026//0949-3964.46.3.217>
- Schödl, A. & Göhring, A. (2017). FALKO-P: Fachspezifische Lehrerkompetenzen im Fach Physik. Entwicklung und Validierung eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Physiklehrkräften. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 201–244). Waxmann.
- Schönbrodt, F. D. & Wagenmakers, E.-J. (2017). Bayes factor design analysis: Planning for compelling evidence. *Psychon Bull Rev*, 25, 128–142. <https://doi.org/10.3758/s13423-017-1230-y>
- Schopf, C. & Zwischenbrugger, A. (2015). Verständliche Erklärungen im Wirtschaftsunterricht: Eine Heuristik basierend auf dem Verständnis der Fachdidaktiker/innen des Wiener Lehrstuhls für Wirtschaftspädagogik. *Zeitschrift für ökonomische Bildung*, 3, 1–31.
- Schreiber, W. (2005). Historisches Lernen und Lebenswelt. In B. Weber, B. Stalla & P. Merkel-Trinkwalder (Hrsg.), *Phänomenologische Dimensionen der Bildungsanthropologie: Interdisziplinäre Forschungsbeiträge im Fokus ethischer Verantwortlichkeit* (S. 311–323). Roderer Verlag.
- Schröter, G. (2021). Spielbewegungen im Instrumentalunterricht mit Hilfe somatischer Verfahren erlernen. In B. Busch (Hrsg.), *Grundwissen Instrumentalpädagogik: Ein Wegweiser für Studium und Beruf* (2. Aufl., S. 63–74). Breitkopf & Härtel.
- Schulten, M. L. (1980). Schülerwünsche zu Unterrichtsmethoden im Musikunterricht. In K.-E. Behne (Hrsg.), *Einzeluntersuchungen. Referate der AMPF-Jahrestagung 1979 in Hamburg* (S. 96–111). Laaber-Verlag.
- Schwab, G. (2009). Schülerinitiierte Erklärprozesse im Fremdsprachenunterricht einer Hauptschulklasse. In J. Spreckels (Hrsg.), *Erklären im Kontext. Neue Perspektiven aus der Gesprächs- und Unterrichtsforschung* (S. 49–65). Schneider Verlag Hohengehren.
- Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. & Wendt, H. (Hrsg.) (2020). *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann. <https://doi.org/10.31244/9783830993193>
- Schworm, S. & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, 46(4), 426–445. <https://doi.org/10.1016/j.compedu.2004.08.011>
- Seedhouse, P. (2009). How language teachers explain to students what they are supposed to do. In J. Spreckels (Hrsg.), *Erklären im Kontext. Neue Perspektiven aus der Gesprächs- und Unterrichtsforschung* (S. 66–80). Schneider Verlag Hohengehren.
- Seidel, T. (2014). Lehrerhandeln im Unterricht. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. und erw. Aufl., S. 781–806). Waxmann.
- Seidel, T., Blomberg, G. & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 296–306). Beltz. <https://doi.org/10.25656/01:3438>
- Seidel, T. & Thiel, F. (2017). Standards und Trends der videobasierten Lehr-Lernforschung. *Zeitschrift für Erziehungswissenschaften*, 20, 1–21. <https://doi.org/10.1007/s11618-017-0726-6>

- Seifried, J. & Wuttke, E. (2017). Der Einsatz von Videovignetten in der wirtschaftspädagogischen Forschung: Messung und Förderung von fachwissenschaftlichen und fachdidaktischen Kompetenzen angehender Lehrpersonen. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung und Professionalität pädagogischen Personals* (S. 303–322). Springer. https://doi.org/10.1007/978-3-658-07274-2_16
- Selting, M., Auer, P., Barden, B., Bergmann, J., Couper-Kuhlen, E., Günther, S., Quasthoff, U., Meier, C., Schlobinski, P. & Uhmann, S. (1998). Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte*, 173, 91–122.
- Sherin, M.G. (2001). Professional vision of classroom events. In T. Wood, B. Nelson & J. Warfield (Hrsg.), *Beyond classical pedagogy: Teaching elementary school mathematics* (S. 75–93). Lawrence Erlbaum Associates.
- Sherin, M. G. & van Es, A. (2008). Effects of Video Club Participation on Teachers' Professional Vision. *Journal of Teacher Education*, 60 (1), 20–37. <https://doi.org/10.1177/0022487108328155>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Sprau, K. (2017). Strukturierte Kompetenz? Inhalte des LehrplanPLUS (Gymnasium) aus Sicht der Musiktheorie. In B. Hofmann (Hrsg.), *Plan | mä | ßig. Schulmusik unter den Vorzeichen von Bildungsstandards und Kompetenzorientierung. Beiträge zu den Tagen der Bayerischen Schulmusik 2016* (S. 117–132). Helbling.
- Spreckels, J. (2009a). Erklären im Kontext: neue Perspektiven. In J. Spreckels (Hrsg.), *Erklären im Kontext – Neue Perspektiven aus der Gesprächs- und Unterrichtsforschung*. Schneider Verlag Hohengehren.
- Spreckels, J. (2009b). Mündliches Erklären im Deutschunterricht. In M. Krelle & C. Spiegel (Hrsg.), *Sprechen und Kommunizieren* (S. 117–138). Schneider Verlag Hohengehren.
- Spsychiger, M. (2015). Lernpsychologische Perspektiven für eine grundschulspezifische Musikdidaktik. In M. Fuchs (Hrsg.), *Musikdidaktik Grundschule. Theoretische Grundlagen und Praxisvorschläge* (S. 50–71). Helbling.
- Spsychiger, M. (2019). Kreis, koordinativer Raum und pädagogisches Halten im (Musik-)Unterricht: Ausführungen zum Konstrukt der Koordination. In B. Hofmann & G. Puffer (Hrsg.), *Macht Musik. Beiträge zu den Tagen der Bayerischen Schulmusik 2018* (S. 11–29). Helbling.
- Stadtler-Elmer, S. (2000). *Spiel und Nachahmung. Über die Entwicklung der elementaren musikalischen Aktivitäten*. Breitkopf & Härtel.
- Stark, R., Gruber, H., Hinkofer, L., Mandl, H. & Renkl, A. (2001). Entwicklung und Optimierung eines beispielbasierten Instruktionsansatzes zur Überwindung von Problemen der Wissensanwendung in der kaufmännischen Erstausbildung. In K. Beck & V. Krumm (Hrsg.), *Lehren und Lernen in der beruflichen Erstausbildung* (S. 369–387). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-10645-6_17
- Steffen-Witteck, M. (2021). Musik – Bewegung – Tanz. In W. Jank (Hrsg.), *MusikDidaktik* (S. 246–254). Cornelsen Scriptor.
- Stein, M. K. & Kucan, L. (2010). *Instructional Explanations in the Disciplines*. Springer. <https://doi.org/10.1007/978-1-4419-0594-9>
- Stibi, S. (2018). *Tanzimprovisation anleiten. Eine beschreibende Systematik multimodaler Instruktionsformen*. [Dissertation] Universität Augsburg. <https://opus.bibliothek.uni-augsburg.de/opus4/38577> [07.08.2022].
- Stibi, S. (2019). „Ihr seid mit dem Körper die dritte Stimme“. Musikalische Facetten von Instruktionen zur Tanzimprovisation. In G. Puffer, A. Becker, F. Körndle & K. Sprau (Hrsg.), *Musik – Pädagogik – Professionalität. Festschrift für Bernhard Hofmann zum 60. Geburtstag* (S. 245–261). Helbling.

- Strübing, J., Hirschauer, S., Ayaß, R., Krähnke, U. & Scheffer, T. (2018). Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. *Zeitschrift für Soziologie*, 47(2), 83–100. <https://doi.org/10.1515/zfsoz-2018-1006>
- Sumfleth, E., Kobow, I., Özcan, N. & Walpuski, M. (2013). Fachkommunikation im Chemieunterricht. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach: Sprachlichkeit und fachliches Lernen* (S. 255–276). Waxmann.
- Tenorth, H.-E. (2004). Bildungsstandards und Kerncurriculum. Systematischer Kontext, bildungstheoretische Probleme. *Zeitschrift für Pädagogik*, 50, 650–661.
- Tepner, O. & Dollny, S. (2014). Entwicklung eines Testverfahrens zur Analyse fachdidaktischen Wissens. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftlichen Forschung* (S. 311–323). Springer. https://doi.org/10.1007/978-3-642-37827-0_25
- Terhart, E. (2002). *Standards für die Lehrerbildung: eine Expertise für die Kultusministerkonferenz*. ZKL. https://doi.org/10.1007/978-3-322-80934-6_15
- Terhart, E. (2009). *Didaktik – Eine Einführung*. Reclam.
- Terhart, E. (2011). Lehrerberuf und Professionalität. Gewandeltes Begriffsverständnis – neue Herausforderungen. *Zeitschrift für Pädagogik*, 57 (Beiheft), 202–224.
- Terhart, E. (2013). *Erziehungswissenschaft und Lehrerbildung*. Waxmann
- Terhart, E., Bennewitz, H. & Rothland, M. (Hrsg.) (2014). *Handbuch der Forschung zum Lehrerberuf*. Waxmann.
- Thim-Mabrey, C. & Lindl, A. (i. Vorb.). Erklärqualität aus sprachwissenschaftlicher Sicht. Die Wahrnehmung der sprachlichen Verständlichkeit von Erklärvideos aus elf verschiedenen Fächern. In A. Schilcher, A. Lindl, S. Krauss & S. Hilbert (Hrsg.), *Fachspezifische Lehrerkompetenzen im Erklären*. Beltz Juventa.
- Tillmann, K.-J. (2014). Konzepte der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarb. und erw. Aufl., S. 308–316). Waxmann.
- Trautmann, M. & Wischer, B. (2016). Lehren. Überlegungen zu einem nur wenig benutzten Begriff. In A. Feindt, W. Herget, M. Trautmann, B. Wischer & K. Zierer (Hrsg.), *Lehren. Friedrich Jahresheft XXXIV 2016* (S. 4–7). Friedrich Verlag.
- Valicer, W. F. (1978). Suppressor variables and the semipartial correlation coefficient. *Educational and Psychological Measurement*, 38(4), 953–958. <https://doi.org/10.1177/001316447803800415>
- Venus, D. (1969). *Unterweisung im Musikhören*. Noetzel.
- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M. & Pietsch, M. (2020). Angebots-Nutzungsmodelle der Wirkweise des Unterrichts: eine kritische Auseinandersetzung mit ihrer theoretischen Konzeption. *Zeitschrift für Pädagogik*, 66(Beiheft 1/20), 63–80.
- Vogel, C. (2018). Tanzen/Bewegen als didaktisches Handlungsfeld. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 334–340). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Vogt, J. (2008) Musikbezogene Bildungskompetenz – ein hölzernes Eisen? Anmerkungen zu den Theoretischen Überlegungen zu einem Kompetenzmodell für das Fach Musik. *Zeitschrift für Kritische Musikpädagogik*, 34–41. <http://www.zfkm.org/sonder08-vogt.pdf> [07.08.2022].
- Vogt, J. (2012). Musikalische Bildung – ein lexikalischer Versuch. *Zeitschrift für kritische Musikpädagogik*, 2012, 1–25.
- Vogt, J. (2016). Didaktische Interpretation von Musik. In A. Lehmann-Wermser (Hrsg.), *Musikdidaktische Konzeptionen. Ein Studienbuch* (S. 31–44). Wißner.

- Vogt, R. (2009). Die Organisation von Erklärprozessen im Unterricht. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (2. Aufl., S. 203–225). Stauffenburg Linguistik.
- Voigt-Zimmermann, S. (2017). Auswirkungen der heiseren Stimme von Pädagogen auf die Leistungen von Kindern. In M. Fuchs (Hrsg.), *Die Stimme im pädagogischen Alltag* (S. 37–47). Logos.
- Voss, T., Kunina-Habenicht, O., Gindele, V. & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, 18(2), 187–223. <https://doi.org/10.1007/s11618-015-0626-6>
- Voss, T. & Kunter, M. (2011). Pädagogisch-psychologisches Wissen 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. 193–214). Waxmann. <https://doi.org/10.31244/9783830974338>
- Wagner, A. & Wörn, C. (2011). *Erklären lernen – Mathematik verstehen. Ein Praxisbuch mit Lernangeboten*. Klett.
- Wagner, W., Göllner, R., Helmke, A., Trautwein, U. & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and instruction*, 28, 1–11. <https://doi.org/10.1016/j.learninstruc.2013.03.003>
- Wallbaum, C. (2009). *Produktionsdidaktik im Musikunterricht. Perspektiven zur Gestaltung ästhetischer Erfahrungssituationen* (2. Aufl.). Qucosa.
- Wallbaum, C. (2010) Wenn Musik nur in erfüllter Praxis erscheint. Ästhetische und kulturelle Kriterien für die Untersuchung und Gestaltung von Musikunterricht. In C. Wallbaum (Hrsg.), *Perspektiven der Musikdidaktik. Drei Schulstunden im Licht der Theorien* (S. 83–122). Olms.
- Weber, R. (2009). Musik begreifen – Musiklehre – Musiktheorie. In S. Helms, R. Schneider & R. Weber (Hrsg.), *Handbuch des Musikunterrichts: Primarstufe* (4. Aufl., S. 265–276). Bosse.
- Weidner, V. (2012). ‚Die‘ Musiktheorie ‚der‘ Musikpädagogik. Systemtheoretische Beobachtungen. In J. Knigge & A. Niessen (Hrsg.), *Musikpädagogisches Handeln. Begriffe, Erscheinungsformen, politische Dimensionen* (S. 300–315). Die Blaue Eule.
- Weidner, V. (2015). *Musikpädagogik und Musiktheorie. Systemtheoretische Beobachtungen einer problematischen Beziehung*. Waxmann.
- Weidner, V. (2018a). Gesprächsführung im Musikunterricht. Wege zum gelingenden Diskurs. *Mip Journal*, 51, 6–10.
- Weidner, V. (2018b). Musikhören als didaktisches Handlungsfeld. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 315–318). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Weidner, V. (2018c). Reflexion als didaktisches Handlungsfeld. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (S. 348–354). Waxmann/UTB. <https://doi.org/10.36198/9783838550404>
- Weinert, F. E. (1997). Notwendige Methodenvielfalt. Unterschiedliche Lernfähigkeit der Schüler erfordern variable Unterrichtsmethoden des Lehrers. *Friedrich Jahresheft*, 1997, 50–52. Friedrich Verlag.
- Weinert, F. E. (1999). *Konzepte der Kompetenz*. OECD.
- Weinert, F. E. (2001). Concept of competence: a conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Hogrefe & Huber Publishers.
- Wellenreuther, M. (2013). *Lehren und Lernen – aber wie? Empirisch-experimentelle Forschungen zum Lehren und Lernen im Unterricht* (8., korr. u. überarb. Aufl.). Schneider Verlag Hohengehren.

- Wisniewski, B., Zierer, K., Dresel, M. & Daumiller, M. (2020). Obtaining secondary students' perceptions of instructional quality: Two-level structure and measurement invariance. *Learning and Instruction*, 66. <https://doi.org/10.1016/j.learninstruc.2020.101303>
- Wisniewski, B., Zierer, K. & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis on Educational Feedback Research. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.03087>
- Wittwer, J. & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, 43(1), 49–64. <https://doi.org/10.1080/00461520701756420>
- Wörn, C. (2014). *Unterrichtliche Erklärsituationen. Eine empirische Studie zum Lehrerhandeln und zur Kommunikation im Mathematikunterricht der Sekundarstufe I*. Verlag Dr. Kovač.
- Wragg, E. & Wood, E. (1984). Pupil Appraisals of Teaching. In Wragg, E. (Hrsg.), *Classroom Teaching Skills* (S. 79–96). Croom Helm. <https://doi.org/10.4324/9780203325445>
- Zimmer, B. (2002). Das Konzept der Lebenswelt. Fluchtpunkt oder Verheißung für die Musikpädagogik? In R.-D. Kraemer (Hrsg.), *Multimedia als Gegenstand musikpädagogischer Forschung* (S. 169–191). Die Blaue Eule.