

*Mario Frei, Gabriele Puffer, Sven Hilbert & Bernhard Hofmann*

## **Merkmale guten Erklärens im Musikunterricht**

Eine Frage der Perspektive?

*Characteristics of “Good Explaining” in Music Lessons.  
A Matter of Perspective?*

*Explaining is one of the central activities of a teacher in the classroom. Although initial theoretical models are available on the question of which characteristics constitute good explaining in music lessons, empirical evidence on this is still missing. This paper presents key findings from the FALKE-q-Mu study. Assessments of selected characteristics of instructional explanation from different perspectives were analyzed.*

### **1. Einleitung**

Unterrichtliches Erklären wird im Folgenden verstanden als „komplexer, vorbereiteter oder sich situativ ergebender sowie interaktiver Kommunikationsprozess, der auf eine Fähig- beziehungsweise Fertigkeitentwicklung sowie die Initiierung eines Verstehensprozesses bei Adressatinnen und Adressaten abzielt“ (Lindl et al., 2019, S. 131). Es zählt zu den Kerntätigkeiten einer Lehrkraft und hat große Bedeutung für qualitätsvolle Unterrichtsgestaltung (Leinhardt, 2001; Pauli, 2015). Einschlägige Zuschreibungen im Kontext der Lehrkräftebildung (z. B. Cramer et al., 2020; Fraefel & Scheidig, 2018) und die Zunahme an empirischen Studien in verschiedenen Fachdidaktiken (z. B. Findeisen, 2017; Neumeister, 2011) machen deutlich, dass das Thema in Unterrichtsforschung und Lehrer\*innenbildung zunehmend Interesse auf sich gezogen hat. Dabei zeigte sich, dass fachlichen Spezifika große Bedeutung zukommt (Leinhardt, 2001). So liegt es nahe, sich mit Erklären auch unter Perspektive des Fachs Musik zu befassen und dadurch einem musikpädagogischen Desiderat zu begegnen.

Im Rahmen der Studie *FALKE-q-Mu*<sup>1</sup> wurden Beurteilungen verschiedener Statusgruppen (z. B. Schüler\*innen und Lehrkräfte) zu ausgewählten Merkmalen unterrichtlichen Erklärens im Musikunterricht erhoben und analysiert. Als Teil des transdisziplinären Forschungsprojekts *FALKE-q* (**F**achspezifische **L**ehrer**k**ompetenzen im **E**rklären – **Q**ualität; Schilcher et al., 2021) hat sie Merkmale guten Erklärens in elf Fächern und unter Beteiligung von insgesamt 14 Disziplinen untersucht.<sup>2</sup> Das Projekt verfolgte einen Ansatz, der generische und fachspezifische Aspekte berücksichtigt und entsprechenden Postulaten aus Unterrichts- und Lehrkräftenprofessionsforschung Rechnung trägt (vgl. Klieme, 2006; Praetorius et al., 2020). Der Beitrag stellt das Forschungsdesign vor und berichtet ausgewählte Ergebnisse.

## 2. Theoretischer Hintergrund

Im Rahmen des Projekts wurden auf Basis einer fachübergreifenden Literaturrecherche vier Qualitätsmerkmale<sup>3</sup> für unterrichtliches Erklären identifiziert: Adressatenorientierung, Strukturiertheit, Sprachliche Verständlichkeit sowie Sprech- und Körperausdruck (Schilcher et al., 2021). Dass diese vier Merkmale über alle beteiligten Fächer hinweg signifikante Prädiktoren für *gutes Erklären* darstellen, konnte auch empirisch gezeigt werden (Lindl et al., 2019). Im Musikunterricht spielen akustische Phänomene und musikpraktische Aktivitäten eine große Rolle, was eine fachspezifische Ergänzung der generischen Rahmung notwendig machte. Dazu wurde das Konzept externer Repräsentation herangezogen. „Repräsentation“ wird dabei nach Schnotz und Bannert (1999) verstan-

- 
- 1 *FALKE-q-Mu* gehört zur Maßnahme *FALKE-q* und ist Teil des Projekts *KOLEG*. *KOLEG* wurde im Rahmen der gemeinsamen Qualitätsoffensive Lehrerbildung von Bund und Ländern mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1512 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor\*innen.
  - 2 In alphabetischer Reihenfolge: Bildende Kunst und Ästhetische Erziehung (Matthias Weich, Birgit Eiglsperger), Biologie (Christina Ehras, Arne Dittmer), Chemie (Michael Elmer, Oliver Tepner), Deutsch (Lisa Gaier, Anita Schilcher), Englisch (Maria Gastl-Pischetsrieder, Petra Kirchhoff), Evangelische Religion (Renate Murmann, Michael Fricke), Geschichte (Anna-Maria Ruck, Josef Memminger), Grundschulpädagogik (Katharina Asen-Molz, Astrid Rank), Mathematik (Simone Röhl, Stefan Krauss), Musik (Mario Frei, Gabriele Puffer, Bernhard Hofmann) und Physik (Jana Heinze, Karsten Rincke).
  - 3 Die Bezeichnungen *Qualitätsmerkmale* und *Merkmale guten Erklärens* werden im vorliegenden Text synonym verwendet, beziehen sich jedoch im Sinne der Klassifikation von Unterrichtsqualität nach Berliner (2005) jeweils auf normative Prinzipien guten Erklärens. Auf Basis der vorliegenden Daten kann keine Aussage hinsichtlich der Wirksamkeit von Erklären getroffen werden.

den als „Objekt oder ein Ereignis, das für etwas anderes steht, es re-präsentiert“ (S. 5). Traditionsreiche Komponenten unterrichtlichen Erklärens im Fach Musik sind visuelle und verbale Formen der Repräsentation, beispielsweise grafische Veranschaulichungen von Tonhöhenverläufen (z. B. Richter, 2007, S. 109–112) oder die Verwendung von Analogien oder Metaphern (z. B. Oberschmidt, 2011). Im Musikunterricht sind darüber hinaus akustische Formen der Repräsentation bedeutsam, wie etwa Klang- oder Hörbeispiele, im Kontext praktischen Musizierens können auch taktile Repräsentationen eine Rolle spielen (Frei et al., 2022). Repräsentationsformen lassen sich kombinieren und zu je verschiedenen Erklärmodi zusammenfügen (Findeisen, 2017).

Als relevant für die fachspezifische theoretische Modellierung, die der Studie FALKE-q-Mu zugrunde liegt, erwies sich eine Abgrenzung des (unterrichtlichen) Erklärens vom (unterrichtlichen) Instruieren: *Erklären* zielt auf den Auf- und Ausbau von Wissen, während *Instruieren* zu Handlungsvollzügen befähigen soll (Hohenstein, 2009). Mithin sind im Musikunterricht Instruktionshandlungen der Lehrkraft wichtig, insbesondere beim Initiieren und Anleiten von Musizierprozessen. Auf der Ebene beobachtbaren Handelns von Lehrkräften können sich Instruktionen und Erklärungen ähneln und/oder überlappen, dementsprechend werden dieselben Qualitätsmerkmale vermutet (Frei et al., 2022). Mit Blick auf die jeweils unterschiedlichen Zielsetzungen lassen sich Erklären und Instruieren jedoch analytisch trennen. Wenn beispielsweise eine Musiklehrkraft etwas vorsingt, um stimmphysiologische Vorgänge klanglich zu veranschaulichen, fällt das unter *Erklären*, wenn sie damit die Schüler\*innen zum wohlklingenden Nachsingen befähigen möchte, unter *Instruieren*. Die beobachtbaren Handlungen gleichen sich, können aber anhand des verfolgten Zwecks unterschieden werden (Frei et al., 2022). Abbildung 1 veranschaulicht die vorgenommene Modellierung (siehe Frei et al., 2022, S. 205).

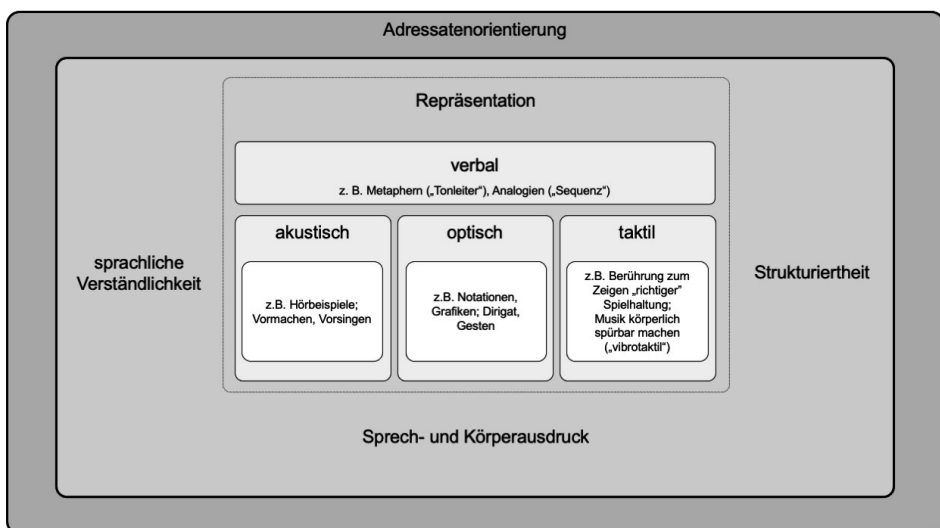


Abbildung 1: Modell guten Erklärens im Musikunterricht (Frei et al., 2022, S. 205)

Diese Konzeptualisierung unterrichtlichen Erklärens entstand mit Blick auf das unterrichtliche Erklären basaler musiktheoretischer Inhalte (z. B. Notenwerte, Taktarten) und damit auf einen ebenso *klassischen* wie umstrittenen Ausschnitt schulischen Musikunterrichts (Frei et al., 2022).

Eine Durchsicht historischer und aktueller schulischer Lehrpläne für das Fach Musik zeigt, dass sich entsprechende Inhalte als zeitlich überdauernde, robuste Komponenten erweisen (z. B. Kemmelmeyer & Nykrin, 2004). Über deren grundsätzliche Sinnhaftigkeit sowie über einschlägige Vermittlungsformen wird intensiv diskutiert (z. B. Weidner, 2015), ohne dass dabei bisher der Fokus auf der Rolle des Erklärens lag. Im Rahmen einer transdisziplinär verorteten Studie boten sich solche musiktheoretischen Unterrichtsinhalte dennoch in besonderer Weise an: Repräsentationsformen und Erklärmodi können hier gut einzeln und kontrolliert untersucht werden (vgl. dazu Frei et al., 2022). Der Fokus auf kognitive Inhalte ermöglicht zudem eine gute Vergleichbarkeit mit dem Erklären in anderen Fächern.<sup>4</sup>

### 3. Forschungsfragen

Die Studie geht von der Annahme aus, dass die Beurteilung der Qualität unterrichtlichen Erklärens grundsätzlich perspektivenspezifisch ist. Diese Vermutung stützt sich auf Befunde aus dem Diskurs um Unterrichtsqualität: Akteur\*innen im Kontext schulischen Unterrichts, wie Schüler\*innen, Lehrkräfte oder externe Beobachtende, können in der Beurteilung von Unterrichtsqualität gleiche, ähnliche oder auch divergente Sichtweisen einnehmen. Somit können sich Qualitätsmerkmale zeigen, die unter mehreren oder unter nur einer Perspektive relevant sind. Um ein möglichst vollständiges Bild zu erhalten, ist folglich ein mehrperspektivischer und zudem perspektivensensibler Ansatz sinnvoll (Kunter & Baumert, 2006; Kleickmann et al., 2019). Die vorliegende Studie untersuchte daher unter anderem, ob sich eine entsprechend differenzierte Validität der postulierten Qualitätsmerkmale mit Blick auf beispielhafte Erklärungen basaler musiktheoretischer Inhalte im Musikunterricht ausmachen lässt. Daraus ergaben sich folgende Forschungsfragen:

*F1a) Welche der theoretisch konzeptualisierten Merkmale guten Erklärens sind für die Bewertung von Erklärqualität relevant?*

---

4 Um mögliche Missverständnisse auszuschließen, sei klargestellt, dass diese Studie weder einem musiktheorielastigen Musikunterricht das Wort redet noch einer überholten Musikdidaktik, die deklaratives Wissen vor praktisches Handeln setzt.

*F1b) Inwieweit liegen dabei Unterschiede zwischen verschiedenen Statusgruppen (Schüler\*innen, Studierende und Lehrende<sup>5</sup>) vor?*

Eine weitere Frage leitete sich ab aus der Kombination perspektivenspezifischer Validität mit der oben postulierten Relevanz akustischer Repräsentationen für den Musikunterricht:

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

#### 4. Methode

##### Design

Angehörige verschiedener Statusgruppen bewerteten mit Hilfe eines Onlinefragebogens dieselben sieben Erklärungen (siehe Abb. 2). Um die verschiedenen Sichtweisen zu berücksichtigen, lag der Fragebogen in zwei Versionen vor: Für Schüler\*innen wurden die Items aus Sicht von Adressat\*innen einer Erklärung formuliert (z. B. *Ich konnte der Erklärung mit meinem Wissen gut folgen*), für Studierende und Lehrende aus der Perspektive derer, die Erklärungen geben (z. B. *Schüler\*innen konnten der Erklärung mit ihrem Wissen gut folgen*). Um eine

| Onlinefragebogen<br>(Beurteilung von 7 videografierten Erklärungen einer Musiklehrkraft) |   |   |   |
|--|---|---|---|
| Beurteilung derselben Videos   | 7 Videos<br>à ca. drei Minuten                      | 3 bzw. 4 ausgewählte<br>Videoausschnitte<br>à ca. 30 Sekunden                                     | 7 Videos<br>à ca. drei Minuten  |
| Teilnehmer*innen<br>(Statusgruppen):   | Holistische<br>Bewertung<br>(Schulnote von 1 bis 6) | Kriterienbasierte<br>Beurteilung<br>(Ratingskala von 1 bis 6)                                     | Kriterienbasierte<br>Beurteilung<br>(Ratingskala von 1 bis 6)   |
|  | Schüler*innen<br>-----<br>Studierende<br>Lehrende   | Items zum<br>Sprech- und Körperausdruck<br>sowie zur<br>Persönlichkeitswirkung<br>des Erklärenden | Items zu<br>Einsatz musikbezogener<br>Repräsentationen,<br>Sprachliche Verständlichkeit<br>Strukturiertheit und<br>Adressatenorientierung |
|  | Durchgang 1   |   | Durchgang 2   |

Abbildung 2: Forschungsdesign

5 Die Gruppe der Lehrenden umfasst sowohl bayerische (Seminar-)Lehrkräfte für das Fach Musik als auch Personen, die an der universitären Ausbildung zukünftiger Musiklehrkräfte beteiligt sind.

holistische Bewertung der Erklärqualität und eine Beurteilung entlang ausgewählter Merkmale zu trennen, wurden die Videos in zwei separaten Durchläufen präsentiert (siehe Abb. 2).

### *Erhebungsinstrument*

Als Stimuli im Onlinefragebogen wurden kurze, systematisch variierte Videos verwendet (*scripted videos*, Piwowar et al., 2018; Dauer jeweils ca. 3 Minuten). Um eine praxisnahe Anmutung zu erreichen, zeigen die Videos Erklärhandlungen, wie sie im Musikunterricht vorkommen könnten. Die Filme wurden in einem Musiksaal in Anwesenheit von Schüler\*innen gedreht. In allen Videos agiert dieselbe männliche Lehrkraft. Die Erklärungen erfolgen monologisch, auf eine Interaktion mit Schüler\*innen wurde bewusst verzichtet, um auf das Erklärhandeln der Lehrkraft fokussieren zu können. Der Lehrer erklärte ohne Skript, Erklärthema und Erklärmodus waren vorgegeben.

Tabelle 1: Systematische Variation von Erklärthema und Erklärmodus

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

### *Operationalisierung*

Die Konstruktion der Fragebogenitems unterlag einem mehrschrittigen Entwicklungs- und Pilotierungsprozess (Frei, 2019). Auf Basis einer ausführlichen Literaturrecherche sowie der Bewertung eines ersten Pools an Items durch Expert\*innen aus Schule und Lehrkräftebildung erfolgte die Pilotierung an Mitgliedern aller Statusgruppen unter Verwendung der kognitiven Interviewtech-

nik *Probing* (Bühner, 2021). Für einen ersten Eindruck soll für jedes Qualitätsmerkmal jeweils ein Item als Beispiel dienen (siehe Tab. 2):

Tabelle 2: Beispielitems zu den fünf Qualitätsmerkmalen

| Merkmalsmerkmal (Itemzahl)                       | Itemformulierung   | Perspektiven |
|--|--|--------------|
| Adressatenorientierung (5 SL/3 SuS) <sup>6</sup> | Schüler*innen konnten der Erklärung mit ihrem Wissen gut folgen.   | SL           |
|  | Ich konnte der Erklärung mit meinem Wissen gut folgen.   | SuS          |
| Strukturiertheit (4)                             | Der Lehrer hat die Erklärung Schritt für Schritt aufgebaut.  | SL/SuS       |
| Sprachliche Verständlichkeit (3)                 | Bei manchen Wörtern wissen die Schüler*innen eventuell nicht, was sie bedeuten.  | SL           |
|  | Bei manchen Wörtern habe ich nicht gewusst, was sie bedeuten.  | SuS          |
| Sprech- und Körperausdruck (7)                   | Der Lehrer hatte eine angemessene Sprechgeschwindigkeit.   | SL/SuS       |
| Einsatz musikbezogene Repräsentation (4)         | Für Schüler*innen einer 6. Jahrgangsstufe ist wichtig, ... dass der Lehrer eine Präsentation an der Tafel verwendet hat. | SL           |
|  | ... dass der Lehrer Hörbeispiele auf dem Klavier vorgespielt hat.  |              |
|  | Für mich war wichtig, ... dass der Lehrer eine Präsentation an der Tafel verwendet hat.                                  | SuS          |
|  | ... dass der Lehrer Hörbeispiele auf dem Klavier vorgespielt hat.  |              |

Bem.: SL = Studierende und Lehrende; SuS = Schülerinnen und Schüler.

Des Weiteren wurde mit drei Items<sup>7</sup> nach Beurteilungen zur Lehrperson gefragt (Sympathie, Enthusiasmus und Authentizität). Unter der Sammelbezeichnung *Persönlichkeitswirkung* wurden sie als Kontrollvariablen genutzt, um entsprechende Effekte zu kontrollieren und die Fehlervarianz abschätzen zu können (vgl. Keller et al., 2016; Kunter et al., 2011).

6 Aus inhaltlichen Gründen unterscheidet sich die Itemzahl für die Skala Adressatenorientierung zwischen Schüler\*innen (3 Items) und Studierenden sowie Lehrenden (5 Items). Letztere wurden beispielsweise nach dem Grad an didaktischer Reduktion in der Erklärung gefragt; Schüler\*innen fehlt für eine solche Einordnung die didaktische Kompetenz.

7 Beispiel: „Der Lehrer wirkte in diesem Video auf mich sympathisch“.

### Durchführung

Die Datenerhebung erfolgte in zwei Schritten und verlief für alle Statusgruppen auf dieselbe Weise (siehe Abb. 2). Zunächst bewerteten alle Studienteilnehmer\*innen jeweils sieben Erklärungen zu drei verschiedenen Themen. Dabei waren sie aufgefordert, in Form einer Schulnote von 1 bis 6 (bei Bedarf mit einer Tendenz +/-) eine holistische Bewertung der Qualität der Erklärung abzugeben. Der zweite Schritt bestand aus einer kriteriengeleiteten Beurteilung der theoretisch postulierten Einzelmerkmale von Erklärqualität. Dadurch konnte gewährleistet werden, dass alle Studienteilnehmer\*innen zunächst die Erklärqualität bewerteten, ohne dass sie dabei von den in der Konzeptualisierung als relevant erachteten Qualitätsmerkmalen beeinflusst wurden. Gleichzeitig können im Nachhinein Zusammenhänge zwischen den Globalurteilen und den Urteilen auf Basis der vorgegebenen Qualitätsmerkmale analysiert werden. Die Beurteilung des Merkmals *Sprech- und Körperausdruck* sowie der Kontrollvariable *Persönlichkeitswirkung* erfolgte in Orientierung am *thin-slices*-Prinzip (z. B. Babad, 2007): Gezeigt wurden kurze, ca. 30-sekündige Sequenzen ausgewählter Erklärvideos. Für die Bewertung aller weiteren Qualitätsmerkmale wurden nochmals alle Erklärungen in kompletter Länge angesehen.

### Stichprobe

Die Stichprobe setzt sich zusammen aus 148 Schüler\*innen der Jahrgangsstufe 6 aus Gymnasium und Realschule, 61 Musiklehramtsstudierenden (hauptsächlich Realschule und Gymnasium) sowie 62 Personen, die entweder als Musiklehrkräfte in Gymnasium oder Realschule oder als Fachdidaktiker\*innen in der Universitäts- oder Hochschullehre tätig und dabei zum großen Teil an der Ausbildung von Musiklehrkräften beteiligt sind (siehe Tab. 3).

Tabelle 3: Überblick der Gesamtstichprobe

| Statusgruppe     | N   | Geschlecht<br>w/m/d | Schulart<br>GS/MS/RS/GY | Jahrgangsstufe bzw.<br>Erfahrung M (SD) | Funktion bzw.<br>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           |   |                                    |

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



Die Rekrutierung der insgesamt  $N = 271$  Studienteilnehmer\*innen basierte vielfach auf persönlicher Kontaktaufnahme, weshalb eine nichtzufällige Gelegenheitsstichprobe vorliegt. Die Erhebungen fanden im schulischen Unterricht (Schüler\*innen), in universitären Lehrveranstaltungen (Studierende) oder privat am eigenen Computer (Studierende und Lehrende) statt und erstreckten sich über den Zeitraum Oktober 2017 bis Dezember 2018.

### *Analyseverfahren*

Die Erhebung der quantitativen Daten erfolgte mit der Onlinefragebogen-Software *Unipark*. Im Anschluss wurden mit der Statistiksoftware *R* (*R Core Team*, 2020) zunächst die negativ gepolten Items rekodiert (z. B. das Item zu *Sprachliche Verständlichkeit*, siehe Tab. 2). Die Datenanalyse erfolgte mit deskriptiven und inferenzstatistischen Verfahren.

Zur Beantwortung der ersten Forschungsfrage nach dem Zusammenhang von Globalurteil und den einzelnen Qualitätsmerkmalen unterrichtlichen Erklärens wurde zunächst ein Modell über die gesamte Stichprobe (F1a) und dann getrennt nach den drei Statusgruppen (F1b) je ein Mehrebenenmodell (hierarchisch lineares Regressionsmodell) gerechnet. In dieses flossen die Globalbewertung der Erklärqualität als abhängige Variable und die fünf Qualitätsmerkmale sowie die Kontrollvariable zur *Persönlichkeitswirkung* als Prädiktoren ein. Die Daten sind nach Personen (intraindividueller Faktor) sowie nach Videos (interindividueller Faktor) geschachtelt, um diesen personen- und videobezo-genen Anteil der Varianz aus dem Modell herausrechnen zu können (siehe z. B. Hilbert et al., 2019). Um den Einfluss der *random effects* zu schätzen, wurde die Intraklassenkorrelation (kurz *ICC*) nach Field (2013) ermittelt.<sup>8</sup>

Um zu beantworten, inwieweit die Varianz in den Globalbewertungen durch die Faktoren Erklärmodus und Statusgruppe erklärt werden kann (F2), wurde für jedes Erklärthema eine *Mixed ANOVA* berechnet.<sup>9</sup> Die Globalnote zur Erklärqualität dient als abhängige Variable, die Faktoren Statusgruppe und Erklärmodus gehen als unabhängige Variablen in die Analyse ein.

---

8 Die Analyse erfolgte mit dem R-Paket *lme4* (Bates et al., 2022), die Wahrscheinlichkeitsberechnung eines Fehlers erster Art bei der Signifikanztestung der Regressionsgewichte mit dem R-Paket *lmertest* (Kuznetsova et al., 2017).

9 Die Analyse erfolgte mit dem R-Paket *ez* (Lawrence, 2016). Die notwendige Umstrukturierung der Daten in einen Längsschnitt erfolgte mittels des R-Pakets *multilevel* (Bliese et al., 2022) und dient dazu, die sieben Bewertungen zu den sieben Videos auf eine Variable der Bewertung und eine zweite Faktorvariable hinsichtlich der Messzeitpunkte umzuwandeln. Bei einer *Mixed ANOVA* kann sowohl die Abhängigkeit von Daten (Messwiederholungsfaktor: Erklärmodus) als auch die Unabhängigkeit von Daten (Zwischensubjektfaktor: unterschiedliche Statusgruppen) berücksichtigt werden.

## 5. Ergebnisse

### *Psychometrische Gütekriterien*

Aufgrund standardisierter Instruktionen bei der Bearbeitung des Onlinefragebogens konnte eine objektive Durchführung der Erhebungen für die Gruppe der Schüler\*innen gewährleistet werden. Für die beiden anderen Gruppen war dies aufgrund der überwiegend häuslichen Bearbeitung nicht möglich. Die geschlossenen Itemformate ermöglichten eine objektive Auswertung.

Nach Ausschluss negativ korrelierter Items<sup>10</sup> betragen die Längen der einzelnen Skalen für alle Statusgruppen: *Sprech- und Körperausdruck* (7 Items), *Strukturiertheit* (3 Items), *Sprachliche Verständlichkeit* (3 Items), *Einsatz musikbezogener Repräsentationen* (4 Items) und *Persönlichkeitswirkung* (3 Items). Um den angenommenen Unterschieden der jeweiligen Perspektiven Rechnung zu tragen, werden die internen Konsistenzen im Folgenden getrennt nach Statusgruppen angegeben (Minima- und Maximawerte über alle Videos hinweg).<sup>11</sup> Bei den Schüler\*innen bewegen sich die Werte mit Ausnahme der Skalen *Adressantenorientierung* und *Sprachliche Verständlichkeit* ( $0,32 \leq \alpha \leq 0,68$ ) im Bereich von  $\alpha = 0,66$  (Min, NW\_oa) bis  $\alpha = 0,93$  (Max, TA\_oa) und sind damit nach Field (2013) akzeptabel bis gut. Für die Studierenden können, abgesehen von geringeren Werten bei der Skala zum *Einsatz musikbezogener Repräsentationen* ( $0,35 \leq \alpha \leq 0,72$ ), alle weiteren internen Konsistenzen als akzeptabel bis gut eingestuft werden ( $0,60 \leq \alpha \leq 0,88$ ). Bei der Gruppe der Lehrenden liegen insgesamt über alle Skalen hinweg akzeptable bis gute Werte vor ( $0,55 \leq \alpha \leq 0,94$ ), wobei in einzelnen Videos auch schwächere Schätzer zu finden sind (z. B. *Persönlichkeitswirkung*,  $\alpha = 0,52$ , TL\_oa; *Sprachliche Verständlichkeit*,  $\alpha = 0,48$ , NW\_o). Trotz dieser niedrigen Werte wurden keine Items aus den Skalen ausgeschlossen, um Statusgruppenvergleiche zu ermöglichen. Gleichwohl wird dies bei der Interpretation der Ergebnisse berücksichtigt.

Als erster Indikator für ökologische Validität dienen entsprechende Einschätzungen von Studienteilnehmer\*innen: Über alle Videos hinweg bejahten Studierende und Lehrende die Frage, ob die gezeigte Erklärsituation im schulischen Musikunterricht vorkommen könne ( $M = 2,25$ ,  $SD = 1,18$ ; Skala von 1 = *stimme völlig zu* bis 6 = *stimme überhaupt nicht zu*).

### *Zusammenhänge zwischen Globalurteil und Qualitätsmerkmalen*

Die Frage, inwieweit die Beurteilungen der einzelnen Merkmale mit den holistischen Bewertungen der Erklärqualität zusammenhängen, wird mithilfe von

<sup>10</sup> Zum Ausschlussverfahren siehe ausführlich Frei, i. Vorb.

<sup>11</sup> Eine ausführliche Darstellung aller relevanten Werte (auch Itemtrennschärfen) findet sich bei Frei (i. Vorb.).

Regressionsanalysen beantwortet – sowohl unter Berücksichtigung aller Studienteilnehmer\*innen als auch getrennt nach den drei Statusgruppen. Für eine erste Einordnung dienen die bivariaten Korrelationen zwischen Qualitätsmerkmalen, der Kontrollvariable *Persönlichkeitswirkung* und dem Globalurteil. Diese werden zunächst für die Gesamtstichprobe (siehe Tab. 4) und im Anschluss getrennt für die drei Statusgruppen (siehe Tab. 5 und 6) dargestellt.

Tabelle 4 zeigt, dass zwischen allen Prädiktoren ausschließlich positiv signifikante Korrelationen vorliegen und diese jeweils in ähnlich hohen Werten mit dem Globalurteil korrelieren. Die beiden Performanzmerkmale *Sprech- und Körperausdruck* und *Persönlichkeitswirkung* korrelieren erwartungsgemäß stark miteinander; mittlere Korrelationen bestehen zwischen *Strukturiiertheit* und den Merkmalen *Adressatenorientierung* sowie *Einsatz musikbezogener Repräsentationen*. Einzelne Verschiebungen und Schwerpunkte werden bei den nach Statusgruppen getrennten Korrelationskoeffizienten deutlich.

Tabelle 4: Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für alle Studienteilnehmer\*innen

| Teil-konstrukt                       | Strukturiert-heit | Adressaten-orientierung | Sprech- und Körperaus-druck | Sprachliche Verständ-lichkeit | Persönlich-keitswirkung | Einsatz musikbez. Repräsen-tationen |
|--------------------------------------|-------------------|-------------------------|-----------------------------|-------------------------------|-------------------------|-------------------------------------|
| Strukturiert-heit                    |                   |                         |                             |                               |                         |                                     |
| Adressaten-orientierung              | 0,56**            |                         |                             |                               |                         |                                     |
| Sprech- und Körperaus-druck          | 0,28**            | 0,20**                  |                             |                               |                         |                                     |
| Sprachliche Verständ-lichkeit        | 0,32**            | 0,21**                  | 0,18**                      |                               |                         |                                     |
| Persönlich-keitswirkung              | 0,28**            | 0,27**                  | 0,70**                      | 0,17**                        |                         |                                     |
| Einsatz mu-sikbez. Repräsen-tationen | 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; \*:  $p \leq 0,05$ ; \*\*:  $p \leq 0,01$ .

Aus Tabelle 5 gehen die Werte der Schüler\*innen hervor. Im Vergleich zur Ebene der Gesamtstichprobe fällt zum einen auf, dass *Adressatenorientierung* schwächer mit dem Globalurteil korreliert. Zum anderen liegen höhere Korrelationen zwischen *Strukturiiertheit* und dem *Einsatz musikbezogener Repräsentationen*

vor. Stabil bleibt der hohe Zusammenhang zwischen *Sprech- und Körperausdruck* und *Persönlichkeitswirkung*.

Tabelle 5: Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für die Gruppe der Schüler\*innen

| Teil-konstrukt                     | Strukturiertheit | Adressatenorientierung | Sprech- und Körperausdruck | Sprachliche Verständlichkeit | Persönlichkeitswirkung | Einsatz musikbez. Repräsentationen |
|------------------------------------|------------------|------------------------|----------------------------|------------------------------|------------------------|------------------------------------|
| Strukturiertheit                   |                  |                        |                            |                              |                        |                                    |
| Adressatenorientierung             | 0,49**           |                        |                            |                              |                        |                                    |
| Sprech- und Körperausdruck         | 0,27**           | 0,17**                 |                            |                              |                        |                                    |
| Sprachliche 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; \*:  $p \leq 0,05$ ; \*\*:  $p \leq 0,01$ .

Tabelle 6 berichtet die Korrelationswerte für die Studierenden (weiß hinterlegt) und die Gruppe der Lehrenden (grau hinterlegt). Neben den weiterhin hohen Korrelationen zwischen *Sprech- und Körperausdruck* sowie *Persönlichkeitswirkung* liegen für beide Statusgruppen hohe, signifikant positive Korrelationen zwischen den Merkmalen *Adressatenorientierung*, *Strukturiertheit* und *Sprachliche Verständlichkeit* vor – allesamt Kriterien, die eher auf einer Tiefenstruktur<sup>12</sup> von Erklären anzusiedeln sind. Während bei den Studierenden mit dem Globalurteil vor allem *Adressatenorientierung* und *Strukturiertheit* mittelstark korreliert sind, besteht bei den Lehrenden eine hohe Korrelation insbesondere mit *Adressatenorientierung*.

12 Zur Terminologie *Tiefen- und Oberflächenstrukturen von Unterricht* siehe z. B. Kunter und Trautwein (2013).

Tabelle 6: Interskalenkorrelationen und deren Zusammenhänge mit dem Globalurteil über alle Videos (Pearson Produkt-Moment-Korr.) für die Gruppen der Studierenden und der Lehrenden

| Teil-konstrukt                     | Strukturiertheit | Adressatenorientierung | Sprech- und Körperausdruck | Sprachliche Verständlichkeit | Persönlichkeitswirkung | Einsatz musikbez. Repräsentationen | Globalurteil |
|------------------------------------|------------------|------------------------|----------------------------|------------------------------|------------------------|------------------------------------|--------------|
| Strukturiertheit                   |                  | 0,69**                 | 0,46**                     | 0,64**                       | 0,42**                 | 0,38**                             | 0,32**       |
| Adressatenorientierung             | 0,70**           |                        | 0,38**                     | 0,62**                       | 0,44**                 | 0,40**                             | 0,47**       |
| Sprech- und Körperausdruck         | 0,18*            | 0,12                   |                            | 0,45**                       | 0,70**                 | 0,16*                              | 0,21**       |
| Sprachliche Verständlichkeit       | 0,63**           | 0,60**                 | 0,21**                     |                              | 0,44**                 | 0,25**                             | 0,26**       |
| Persönlichkeitswirkung             | 0,18*            | 0,17*                  | 0,65**                     | 0,24**                       |                        | 0,25**                             | 0,17*        |
| Einsatz musikbez. Repräsentationen | 0,38**           | 0,34**                 | 0,11                       | 0,26**                       | 0,11                   |                                    | 0,29**       |
| Globalurteil                       | 0,46**           | 0,51**                 | 0,15*                      | 0,23**                       | 0,21**                 | 0,25**                             |              |

Bem.: Korrelationswerte oberhalb der Diagonale für Lehrende, unterhalb der Diagonale für Studierende; \*:  $p \leq 0,05$ ; \*\*:  $p \leq 0,01$ .

Deskriptiv zeigen die bivariaten Interskalenkorrelationen erste Tendenzen, dass die Gruppe der Schüler\*innen eher oberflächenstrukturelle Merkmale wie *Sprech- und Körperausdruck* oder den *Einsatz musikbezogener Repräsentationen* als relevant erachtet, während sowohl bei Studierenden wie Lehrenden der Schwerpunkt eher auf tiefenstrukturellen Merkmalen wie *Adressatenorientierung* und *Strukturiertheit* liegt.

Um die Bedeutung der Einzelmerkmale unterrichtlichen Erklärens für die Gesamtqualität der Erklärungen abzuschätzen, wurden hierarchisch lineare Regressionsmodelle gerechnet. Dabei stellt jeweils die Erklärqualität, gemessen durch das Globalurteil, das Kriterium dar, dessen Varianz durch die ausgewählten Merkmale sowie durch die Kontrollvariable als unabhängige Prädiktoren vorhergesagt wird. Zudem wird dabei jeweils die nach Personen und Videos geschachtelte Datenstruktur berücksichtigt. Dass dies für eine unverzerrte Schätzung notwendig ist, zeigt die Intraklassenkorrelation von 39,09% (siehe dazu Hilbert et al., 2019). Das Mehrebenenmodell mit den Werten aller Studien-

teilnehmer\*innen (siehe Tab. 7) zeigt signifikante positive Regressionsgewichte für alle Merkmale, wohingegen die Kontrollvariable *Persönlichkeitswirkung* keine Bedeutung aufweist.

Tabelle 7: Hierarchisch lineares Regressionsmodell mit der abhängigen Variable Globalurteil und unter Berücksichtigung der nach allen Studienteilnehmer\*innen ( $N_{\max} = 271$ ) und den sieben Videos geordneten Datenstruktur

| Gesamtstichprobe ( $N_{\text{Beob}} = 835$ )   |          |             |
|--|----------|-------------|
| ICC  | %        | 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. Repräsentation               | <i>b</i> | 0,12*       |
| <i>R</i> <sup>2</sup> (marginal   konditional) |          | 0,16   0,43 |

Bem.: *N*: Personenanzahl; Anzahl der Beobachtungen ( $N \times 7$  Videos); ICC: Intraklassenkorrelation; *b*: unstandardisierter Regressionskoeffizient; *R*<sup>2</sup>: Determinationskoeffizient; \*:  $p \leq 0,05$ ; \*\*:  $p \leq 0,01$ .

Die aufgeklärte Varianz beläuft sich für die festen Effekte auf 16% und für die festen und zufälligen Effekte<sup>13</sup> auf 43%, was nach Cohen (1988) als zufriedenstellend eingestuft werden kann. Damit kann die Frage nach der Relevanz einzelner Qualitätsmerkmale unterrichtlichen Erklärens im Fach Musik zunächst damit beantwortet werden, dass alle postulierten Merkmale für die Bewertung der Erklärqualität in etwa derselben Größenordnung bedeutsam sind.

### *Qualitätsmerkmale aus unterschiedlichen Perspektiven*

Um auch die Frage nach möglichen Unterschieden zwischen den einzelnen Perspektiven auf unterrichtliches Erklären zu beantworten, wurden drei Regressionsmodelle gerechnet (eines für jede Statusgruppe, siehe Tab. 8). Es liegen unterschiedliche Intraklassenkorrelationen vor, wobei die Gruppe der Lehrenden das stabilste Beurteilungsverhalten zeigt ( $ICC = 50,83\%$ ).

13 Im linearen Modell sind die festen Effekte alle ausgewählten Qualitätsmerkmale sowie die Kontrollvariable *Persönlichkeitswirkung*, die zufälligen Effekte beziehen sich auf den Einfluss der Personen- und Videostruktur. Der Werte der Varianzaufklärung (*R*<sup>2</sup>) gibt jeweils an, wie viel Prozent der Gesamtvarianz der Globalbewertungen von den festen bzw. festen und zufälligen Effekten der Merkmale erklärt werden kann.

Tabelle 8: Übersicht der drei hierarchischen linearen Regressionsmodelle getrennt nach Statusgruppe

|                                |     | Schüler*innen<br>( $N_{\text{Beob.}} = 480$ ) | Studierende<br>( $N_{\text{Beob.}} = 186$ ) | Lehrende<br>( $N_{\text{Beob.}} = 169$ ) |
|--------------------------------|-----|---|---|--|
| ICC                            | %   | 39,52   | 24,70                                       | 50,83                                    |
| Konstante                      | $b$ | 0,58*   | 0,60*                                       | 0,61                                     |
| Strukturiertheit               | $b$ | 0,09  | 0,38**                                      | 0,06                                     |
| Adressatenorientierung         | $b$ | 0,01  | 0,32**                                      | 0,48**                                   |
| Sprech- und Körperausdruck     | $b$ | 0,25**  | 0,07  | 0,01                                     |
| Sprachliche Verständlichkeit   | $b$ | 0,12**  | -0,20*                                      | 0,01                                     |
| Persönlichkeitswirkung         | $b$ | 0,05  | 0,06  | -0,03                                    |
| Einsatz m. Repräsentation      | $b$ | 0,18**  | -0,06                                       | 0,15                                     |
| $R^2$ (marginal   konditional) |     | 0,16   0,39                                   | 0,22   0,35                                 | 0,25   0,39                              |

Bem.:  $N$ : Personenanzahl; Anzahl der Beobachtungen ( $N \times 7$  Videos); ICC: Intraklassenkorrelation;  $b$ : unstandardisierter Regressionskoeffizient;  $R^2$ : Determinationskoeffizient; \*:  $p \leq 0,05$ ; \*\*:  $p \leq 0,01$ .

Hinsichtlich der Bedeutung der einzelnen Merkmale für die Vorhersage des Kriteriums (globale) Erklärqualität fällt einerseits auf, dass kein Prädiktor in allen drei Modellen signifikant wird (hierbei ist die geringere Teststärke im Vergleich zur Gesamtgruppenanalyse zu berücksichtigen), jedoch für jedes Merkmal in mindestens einem linearen Modell ein signifikanter Regressionskoeffizient vorliegt. Andererseits bestätigen sich die oben bereits beschriebenen, statusgruppenspezifischen Tendenzen der Korrelationen zwischen einzelnen Merkmalen und dem Globalurteil (siehe Tab. 4 bis 6): Bei Studierenden und Lehrkräften scheinen eher tiefenstrukturelle Merkmale wie *Adressatenorientierung* (und bei den Studierenden zusätzlich *Strukturiertheit*) bedeutsam für das Gesamturteil zu sein, während die Globalurteile der Schüler\*innen vor allem von der Qualität der Oberflächenstrukturen, wie *Einsatz musikbezogener Repräsentationen* und *Sprech- und Körperausdruck*, abhängig sind. Die Aussagekraft des signifikanten Regressionskoeffizienten beim Merkmal *Sprachliche Verständlichkeit* ist aufgrund der niedrigen internen Konsistenzen der Skala für die Gruppe der Schüler\*innen stark eingeschränkt, das zugehörige Ergebnis in Form eines negativen Koeffizienten bei den Studierenden kontraintuitiv (siehe ausführlich Frei, i. Vorb.).

Die Regressionsmodelle zeigen, dass für die verschiedenen Statusgruppen innerhalb des gegebenen Konstrukts *Erklärqualität* unterschiedliche Aspekte relevant sind. Vor diesem Hintergrund kann die Vermutung aufrechterhalten

werden, dass Erklärqualität einer perspektivenspezifischen Validität unterliegt. Als weiterer Indikator perspektivenspezifischer Validität kann die Varianzaufklärung bei Studierenden und Lehrenden dienen: Sie fällt jeweils höher aus als im gemeinsamen Modell über alle Studienteilnehmer\*innen (siehe Tab. 7).

### *Unterschiede im Globalurteil bezogen auf den Erklärmodus*

Die systematische Variation der Erklärmodi (siehe Kap. 4) ermöglicht eine Auswertung hinsichtlich der Bedeutung optischer und akustischer Repräsentationen. Die Ergebnisse von Varianzanalysen mit gemischtem Design, in denen das Globalurteil als abhängige und die Faktoren Statusgruppe sowie Erklärmodus als unabhängige Variablen einfließen, verweisen ebenfalls auf statusgruppenspezifische Betrachtungsweisen (für Details: siehe Frei, i. Vorb.), verdeutlichen aber auch die große Bedeutung der Erklärmodi: Sowohl bei den Erklärungen zum Thema Taktarten ( $F(1, 259) = 81,07; p < 0,01; \eta^2 = 0,24$ ) als auch zur Dur-Tonleiter ( $F(2, 259) = 108,22; p < 0,01; \eta^2 = 0,29$ ) liegen starke signifikante Effekte für den Faktor Erklärmodus vor. In der Folge zeigen *Bonferroni*-korrigierte *Post-Hoc*-Tests, dass diese Unterschiede auf die Erklärungen zurückzuführen sind, in denen optische Repräsentationen fehlen.

Das heißt, unabhängig von den drei Themen unterscheiden sich die Bewertungen zwischen Erklärungen, in denen eine optische und akustische Form der Repräsentation vorkommen, und solchen, in denen eine akustische Repräsentation fehlt, nicht. Im Gegensatz dazu werden die Erklärungen signifikant schlechter bewertet, in denen zwar eine akustische Repräsentation vorkommt, eine optische hingegen fehlt. Dieser Befund überrascht insbesondere deshalb, weil eine akustische Repräsentation für eine umfängliche Darstellung von Musik als klingendem Phänomen essenziell erscheint (siehe z. B. Hellberg, 2018; Puffer & Hofmann, 2017). Anhand der vorliegenden quantitativen Daten sind keine Aussagen über Gründe möglich. Studierende und Lehrende konnten ihre Bewertung der Erklärqualität in Form offener Antworten im Fragebogen begründen. Deren qualitative Auswertung lässt die Annahme zu, dass aus Sicht von Erklärenden sowohl in den Videos als auch darüber hinaus optische Repräsentationen stärker dazu geeignet sein könnten, (kognitives) *Verstehen* musiktheoretischer Inhalte zu fördern als die verwendeten akustischen Demonstrationen (siehe ausführlich Frei, i. Vorb.).

## 6. Diskussion

Zusammenfassend lässt sich festhalten, dass alle postulierten Merkmale des Modells (siehe Abb. 1) relevant für die Beurteilung von Erklärqualität sind. Gleichzeitig deuten die Ergebnisse darauf hin, dass das Konstrukt *Erklärqualität* vor dem Hintergrund der ausgewählten musiktheoretischen Sachverhalte



im Unterricht einer perspektivenspezifischen Validität unterliegt. Auffällig ist einerseits, dass keines der theoretisch postulierten Einzelmerkmale von Erklärqualität für alle drei Statusgruppen gleichermaßen von Bedeutung zu sein scheint. Andererseits werden Tendenzen erkennbar, wonach Studierende und Lehrende bei ihren Urteilen verstärkt tiefenstrukturelle Merkmale berücksichtigen, während für Schüler\*innen offensichtlich vornehmlich oberflächenstrukturelle Aspekte Einfluss auf die Bewertung der Erklärqualität haben. Diese perspektivenspezifischen Unterschiede stützen Befunde zur Beurteilung der Unterrichtsqualität von Wisniewski et al. (2020), denen zufolge Schüler\*innen insbesondere das Verhalten der Lehrkraft in den Blick nehmen. Dies könnte auch für die Beurteilung von Erklärqualität relevant sein.

Ein überraschender Befund ergab sich hinsichtlich der Bedeutung optischer und akustischer Repräsentationen für die eingeschätzte Erklärqualität, die auf Basis einer systematischen Variation auf Videoebene untersucht wurde. Erklärungen, in denen eine optische Repräsentation zum Einsatz kam, wurden über alle Statusgruppen hinweg deutlich besser bewertet als solche, in denen sie fehlte. Im Gegensatz dazu zeigen sich keine signifikanten Unterschiede zwischen Erklärungen mit und ohne akustische Repräsentationen. Die vorliegenden qualitativen Daten erlauben lediglich erste Rückschlüsse auf mögliche Ursachen dieses Phänomens. Angesichts der unterrichtspraktischen Relevanz des Themas erscheint weitere theoretische und empirische Arbeit geboten.

Limitationen der vorliegenden Studie liegen einerseits in der fehlenden strukturellen Validierung begründet (z. B. in Form eines Strukturgleichungsmodells, das die theoretische Struktur des Modells bzw. der Qualitätsmerkmale von Erklären im Musikunterricht statistisch überprüft), andererseits in der fehlenden externen Validität der Ergebnisse aufgrund der nichtrepräsentativen Stichprobe. Die ökologische Validität der Untersuchung ist aus zwei Gründen beeinträchtigt: Zum einen basieren die empirischen Erkenntnisse bisher nur auf einem kleinen Ausschnitt von Musikunterricht, in dem vorwiegend kognitive Ziele verfolgt werden. Daher können sie keine Relevanz für andere Bereiche beanspruchen, die sich in Zielsetzung und Interaktionsmustern deutlich vom Untersuchten abheben. Zum anderen fehlen aufgrund der Laborsituation (monologische Erklärung einer Lehrkraft) Interaktionen mit Schüler\*innen, die natürlicher Bestandteil von (Musik-)Unterricht sind. Schließlich mangelt es an prädiktiver Validität: Auf Basis der Daten können keine Aussagen über die Wirksamkeit verschiedener Erklärmodi mit Blick auf Lernerträge und Motivation der Schüler\*innen getroffen werden; gleiches gilt für die Bedeutung einzelner Qualitätsmerkmale.

So wäre eine interessante Frage für weitere Forschungsbemühungen, ob tiefenstrukturelle Merkmale eher dem Lernertrag dienen und oberflächenstrukturelle Merkmale vorrangig der Motivation der Schüler\*innen. Ein wichtiger Ertrag zukünftiger Forschung könnte auch in der Beantwortung der Frage liegen, inwieweit die Bedeutung der ausgewählten Qualitätsmerkmale auch für

Erklärungen im Kontext anderer Bereiche von Musikunterricht oder – bisher gänzlich unklar – für instruierendes Unterrichtshandeln von Musiklehrkräften übertragbar ist.

## Literatur

- Babad, E. (2007). Teachers' Nonverbal Behaviour 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.
- Bates, D., Maehler, 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).
- 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>
- 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].
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* (4. korr. und erw. Aufl.). Pearson.
- Cramer, C., König, J., Rothland, M. & Blömeke, S. (2020). *Handbuch Lehrerinnen- und Lehrerbildung*. Klinkhardt/UTB.
- Cohen, J. (1988). *Statistical Power Analysis für the Behavioral Sciences* (2. Aufl.). Lawrence Erlbaum Associates, Publishers.
- 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.
- Fraefel, 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.
- 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. (i. Vorb.). *Erklären im Musikunterricht. Eine Studie zu Qualitätsmerkmalen* (Dissertation).
- 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]* (S. 197–212). Waxmann.
- Hellberg, B. (2018). Zwischen klingenden Rohdaten und sprachlicher Transformation. In C. Moritz & M. Corsten (Hrsg.), *Handbuch Qualitative Videoanalyse* (S. 217–234). Springer.

- 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>.
- Hohenstein, C. (2009). Interkulturelle Aspekte des Erklärens. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (2. Aufl., S. 37–56). Staufenberg.
- Keller, M. M., Hoy, A. W., Goetz, T. & Frenzel, A. C. (2016). Teacher Enthusiasm: Reviewing and Redefining a Complex Construct. *Educ Psychol Rev*, 28, 743–769. <https://doi.org/10.1007/s10648-015-9354-y>
- Kemmelmeyer, K.-J. & Nykrin, R. (2004). *Spielpläne Musik 5/6* (Neue Ausgabe C Bayern) – Schülerbuch. Klett.
- 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.
- Klieme, E. (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einleitung in den Thementeil. *Zeitschrift für Pädagogik*, 52(6), 765–773.
- 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 CO-ACTIV*. Waxmann.
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts*. Brill, Schöningh/UTB. <https://doi.org/10.36198/9783838538952>
- Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82, 1–26.
- Lawrence, M. (2016). *Ez: Easy Analysis and Visualization of Factorial Experiments* [Software]. <http://github.com/mike-lawrence/ez> (R package version 4.4–0).
- 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.
- Lindl, A., Gaier, L., Weich, M., Frei, M., Ehras, C., Gastl-Pischetsrieder, M., Elmer, M., Asen-Molz, 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.
- Neumeister, N. (2011). *(Wie) Wird im Deutschunterricht erklärt? – Wissensvermittelnde Handlungen im Sprachunterricht der Sekundarstufe I*. Dissertation Pädagogische Hochschule Ludwigsburg. <https://phbl-opus.phlb.de/frontdoor/index/index/docId/32>
- Oberschmidt, J. (2011). *Mit Metaphern Wissen schaffen: Erkenntnispotentiale metaphorischen Sprachgebrauchs im Umgang mit Musik*. Wißner.
- Pauli, C. (2015). Einen Sachverhalt erklären. *Pädagogik*, 3, 44–47.

- Piwowar, 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.
- Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B. & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswiss*, 48, 409–446.
- Puffer, G. & Hofmann, B. (2017). FALKE-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.), *FALKE: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 245–289). Waxmann.
- R Core Team (2021). *Base: The R Base Package* [Software]. <https://www.r-project.org> (R package version 4.1.2).
- Richter, C. (2007). Interpretieren (Werkbetrachtung). In N. Heukäuffer (Hrsg.), *Musik Methodik. Handbuch für die Sekundarstufe I und II* (S. 96–127). Cornelsen Scriptor.
- 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., Röhrl, S., Ruck, A.-M., Weich, M., Dittmer, A., Fricke, M., Hofmann, B., ... Thim-Mabrey, C. (2021). FALKE: Experiences from Transdisciplinary Educational Research by Fourteen Disciplines. *Frontiers in Education*, 5. <https://doi.org/10.3389/feduc.2020.579982>
- Schnotz, W. & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen. *Experimental Psychology*, 46(3), 217–236.
- Weidner, V. (2015). *Musikpädagogik und Musiktheorie. Systemtheoretische Beobachtungen einer problematischen Beziehung*. Waxmann.
- 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>

Mario Frei  
 Universität Regensburg  
 Educational Data Science  
 Sedanstraße 1  
 93055 Regensburg  
 mario.frei@ur.de

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

Sven Hilbert  
Universität Regensburg  
Educational Data Science  
Sedanstraße 1  
93055 Regensburg  
sven.hilbert@ur.de

Bernhard Hofmann  
Universität Augsburg  
Lehrstuhl für Musikpädagogik  
Universitätsstr. 26, 86159 Augsburg  
bernhard.hofmann@uni-a.de