

*Christian Harnischmacher, Mascha Otradinski, Johannes Hasselhorn  
& Miriam Knebusch*

## **Besser geht's nicht!**

Eine empirische Studie zur Leistungsentwicklung und Bewertung des Instrumentalspiels im Schulmusikstudium

*It Doesn't get any Better than this! – An Empirical Study on Performance Development and Assessment of Instrumental Playing in Music Teaching Studies*

*Instrumental education is a key component of the attractiveness of studying music teaching at German music universities. Student performance development is influenced by various factors, including subject-related prejudices and varying levels of expertise. Despite the pivotal role of instrumental training, empirical research on this topic is scarce.*

*This study examined students' self-assessments of their instrumental performance development and the influence of prejudice and expertise on their judgements of instrumental performances. Data collection was conducted through an online listening experiment, employing both quantitative and qualitative analyses.*

*Findings indicate that only two thirds of the surveyed students perceived an improvement in their instrumental abilities. However, no significant influence of prejudice on judgements was detected. Expertise level played a significant role in performance assessment, with more expert students providing stricter evaluations. Qualitative analysis revealed that both musical and non-musical criteria influenced evaluations, with teachers typically assessing more deficit-oriented and with more specific criteria than students.*

### **1. Einleitung**

Die Instrumentalausbildung ist nicht selten ein wesentliches Motiv für die Aufnahme eines Schulmusikstudiums, zumal die Ausbildung am Hauptfachinstrument dabei an den meisten Musikhochschulen den Schwerpunkt bildet (Heyer, 2016, S. 209–212). In unserer Hochschulpraxis berichten Studierende mitunter, dass das geforderte instrumentale Leistungsniveau für die Aufnahme an einer

Musikhochschule nach intensiver und langer Vorbereitung einen „Deckeneffekt“ erzeugen kann. Die Studierenden haben nach eigener Einschätzung bereits zu Studienbeginn ihr mögliches Leistungslevel erreicht und können sich bis zum Master kaum signifikant verbessern. Außerdem schränken die Studienanforderungen eines Lehramtsstudiums ein gewohntes Übepensum oft deutlich ein.

Die Instrumentalausbildung im Hauptfach bildet darüber hinaus auch für Studienbewerber\*innen an den meisten Musikhochschulen das Nadelöhr zum Studienbeginn (vgl. Wolf & Kopiez, 2014), was einen wichtigen Baustein des anhaltenden Lehrkräftemangels im Schulfach Musik darstellt. Dieser Mangel ist nicht nur auf den allgemeinen Attraktivitätsverlust des Lehrberufs zurückzuführen, sondern auch auf fachspezifische Ursachen (Bundesverband Musikunterricht, 2023). Der Bundesverband Musikunterricht appelliert in diesem Zusammenhang an die Verantwortung der Ausbildungsinstitutionen von Schulmusiker\*innen. Dringende Reformen betreffen nach dieser Argumentation eine stärker musikpädagogische und weniger künstlerische Ausrichtung des Lehramtsstudiums im Fach Musik. In der Kritik steht insbesondere „die Hürde der häufig nicht am Berufsfeld orientierten Eignungsprüfungen an den Hochschulen“ (Bundesverband Musikunterricht, 2023, S. 4). Im Kalenderjahr 2022 studierten von den Bewerber\*innen für das Lehramt Musik an deutschen Musikhochschulen anschließend tatsächlich nur etwa 29.6% das Studienfach. Von den dort vorhandenen Studienplätzen für Schulmusik blieben ca. 49% unbesetzt.<sup>1</sup>

Diese Beobachtungen aus unserer Hochschulpraxis werfen grundlegende Fragen zur potenziellen Leistungssteigerung am Instrument in einem Schulmusikstudium und deren Einschätzung durch die Beteiligten auf. Neben der Ausbildung einer künstlerischen Lehrer\*innenpersönlichkeit (vgl. Richter, 2018; Roberts, 1991) wird in Studienordnungen mit dem Hauptfachmodul vor allem eine Verbesserung der eigenen Leistungen im Instrumentalspiel in Aussicht gestellt. Eine empirische Prüfung der anvisierten Leistungsverbesserung steht bislang aus. In diesem Beitrag fokussieren wir die Perspektive der Studierenden und gehen den Fragen nach, wie Studierende selbst ihren Fortschritt am Instrument einschätzen und welche Einflussfaktoren und Kriterien bei ihrem Bewertungsverhalten allgemein eine Rolle spielen.

## 2. Theoretischer Hintergrund

Die instrumentale Leistungsentwicklung von Studierenden im Studium der Schulmusik wurde bislang kaum empirisch untersucht bzw. bestenfalls in Alumnibefragungen erwähnt (Clausen & Schäfer-Lembeck, 2017; Meißner & Schwarz, 1999). In der Pilotstudie zum vorliegenden Beitrag von Hofbauer, Otradinski und

---

1 Angaben der AG Schulmusik vom 23.04.2023.

Wehen (2022) berichten die Schulmusikstudierenden vermehrt von einer Stagnation der Leistungsentwicklung im instrumentalen Hauptfach.

In der vorliegenden Studie wollen wir unter anderem praxisnahe Bewertungskriterien eruieren und nicht Kriterien vorgeben, die die ökologische Validität reduzieren könnten. Entsprechend definieren wir instrumentale Leistungen möglichst offen: Leistungen im Instrumentalspiel beschreiben Folgen instrumentalen Lernens, für die Bewertungskriterien aufgestellt werden (müssen).

Die Post-Pisa-Debatte beförderte mit der Orientierung an Kompetenzen und Standards ein pädagogisch motiviertes Verständnis von Leistung und Leistungsentwicklung, das gesellschaftlich vermittelte Ansprüche an Leistung (soziales Konstrukt) und deren subjektive Bedeutsamkeit (subjektgebundene Konstruktion) als zwei Seiten derselben Medaille auffasst (vgl. Geuen, 2009, S. 58). Entsprechend berücksichtigt Niessen (2020) in einer Situationsanalyse zu Leistungen im Musikunterricht vor allem die Perspektiven der Schüler\*innen. Schulische Leistungen sind demnach weniger eine Frage von Lernergebnissen, sondern von einem demonstrativ leistungsorientierten Unterrichtsverhalten.

In einer stärker forschungsorientierten, operationalisierenden Sichtweise auf Leistungen im Musikunterricht hat sich in musikpädagogischen Studien neben Kompetenztests vor allem das Konstrukt des Kompetenzerlebens empirisch bewährt (Harnischmacher et al., 2021). Im Vergleich dazu erweisen sich Zensuren im Musikunterricht regelmäßig nicht als Prädiktor für musikbezogene Kompetenzen (z.B. Harnischmacher & Knigge, 2017). Auch für die Benotungen der Leistungen im Hauptfachinstrument und der Musiktheorie/Gehörbildung in Eignungsprüfungen an Musikhochschulen konnten Wolf und Kopiez (2014) in einer Längsschnittstudie keinerlei prognostische Evidenz für die weiteren Leistungsbeurteilungen im Verlauf des Bachelorstudiums finden. Unstrittig bleibt derweil, dass die Wahrnehmung und Bewertung der eigenen Leistungen zu den wichtigsten Prädiktoren für eine konstante Entwicklung des Instrumentalspiels oder dessen Abbruch gehören (Switlick & Bullerjahn, 1999). Entsprechend untersuchen wir die Leistungsentwicklung im Instrumentalspiel aus Sicht der Studierenden.

Leistung ist in pädagogischen Kontexten untrennbar mit Bewertungen und Differenzprozessen verbunden (Ricken, 2014). Im Weiteren verstehen wir unter einer Bewertung eine Form des Urteilens. Im Gegensatz zu einer bloßen Meinung sollten Urteile bis zu einem bestimmten Grad auch begründbar sein. Diese Urteile sind in vielfacher Hinsicht erfahrungsbasiert (vgl. Hasselhorn & Wolf, 2018). Zwei in musikbezogenen Bewertungssituationen oft hervorstechende, erfahrungsbezogene Aspekte des Urteils sind *Vorurteile* und *Expertise*.

Bewertungsbeeinflussende Vorurteile in musikbezogenen Situationen können dabei als ein Querschnitt üblicher und aus der pädagogischen Psychologie hinlänglich bekannter Urteilsfehler verstanden werden (z.B. Halo-, Hof- oder Härteeffekt, Hasselhorn & Wolf, 2018). Eine der ersten Studien zum Einfluss von Vorurteilen auf die Bewertung von Leistungen am Instrument hat Klaus-Ernst

Behne (1990) vorgelegt. Dabei wurden Studierende um Bewertungen von Videoaufzeichnungen einer Klaversonate mit verschiedenen Pianist\*innen gebeten. Die Videobeispiele wurden nach Gestik und Erscheinung der Akteur\*innen anders bewertet, obwohl dabei stets die gleiche Audioaufnahme zu hören war. Vergleichbares lässt sich für den Musikunterricht beobachten. In einer experimentellen Studie sollten Schüler\*innen kurze Videobeispiele aus drei verschiedenen Unterrichtsstunden im Fach Musik bewerten (Harnischmacher & Hofbauer, 2013). Die Unterrichtsbeispiele wurden signifikant besser bewertet, wenn die Schüler\*innen von einem Musikunterricht an einem Gymnasium, von einer Fachleiterin für Musik und einem Musiklehrer ausgingen, anstatt von einer Hauptschule, einer fachfremden Lehrerin und einem Referendar. Bewertungen in Musik können also durch Vorurteile determiniert sein. Wir untersuchen in der vorliegenden Studie daher, ob das auch für studienfachbezogene Vorurteile bei der Bewertung instrumentaler Leistungen zutrifft.

Von erfahrungsbasierten Vorurteilen ist eine (ebenfalls erfahrungsbasierte) Expertise abzugrenzen. Expertise wird Personen zugeschrieben, die zumeist aufgrund langjähriger und intensiver Übung (z. B. deliberate practice) in einer bestimmten Domäne dauerhaft hervorragende Leistungen zeigen (Hasselhorn & Knigge, 2018). Expert\*innen haben beispielsweise im Laufe der Jahre ein extensiveres Übepensum absolviert, benutzen effektivere Strategien beim Vom-Blatt-Spielen oder beim Üben. Androustos und Humphrey (2010) untersuchten in einem solchen Zusammenhang erstmals die Beobachtungskompetenz von Musikstudierenden. Mit zunehmender Unterrichtserfahrung reflektieren und kommentieren die Musikstudierenden die dargebotenen Unterrichtsvideos intensiver als Studenten mit weniger Erfahrung. Inwiefern sich Bewertungsunterschiede auch für die Expertise im Instrumentalspiel feststellen lässt, ist daher ebenfalls Teil dieses Beitrags.

### 3. Fragestellung, Methode und Stichprobe

Wir beschränken uns aus pragmatischen Gründen beim Instrumentalspiel auf das Hauptfach von Studierenden an Deutschen Musikhochschulen. Dabei gehen wir folgenden forschungsleitenden Fragen nach:

1. Wie wird die eigene Leistungsentwicklung im Instrumentalspiel von Studierenden der Schulmusik im Studienverlauf eingeschätzt (F1)?
2. Welchen Einfluss haben studienfachbezogene Vorurteile auf die Bewertung instrumentaler Leistungen (F2)?
3. Welchen Einfluss hat die Expertise im Instrumentalspiel auf die Bewertung instrumentaler Leistungen (F3)?
4. Welche qualitativen Merkmale spielen bei der Leistungsbewertung des Instrumentalspiels eine Rolle (F4)?

Die erste Forschungsfrage (F1) zielt dabei auf Selbstauskünfte der Befragten ab. Die weiteren Fragen (F2–F4) verweisen auf Verhaltensbeobachtungen und die Variation von Treatments, nachfolgend beschrieben als Hörbeispiele. Die Datenerhebung wurde in Form einer Onlinebefragung realisiert.

Der erste Teil bestand aus einem Fragebogen, in dessen Verlauf die Probandinnen und Probanden demographische Angaben machen und Fragen zu ihrem instrumentalen Werdegang beantworteten. U.a. wurde die Selbsteinschätzung der Entwicklung des eigenen Instrumentalspiels im Studium mit folgender Frage und Antwortmöglichkeit erhoben:

Wie sehr hat sich Ihr Instrumentalspiel im Hauptinstrument durch das Studium verbessert?

gar nicht      o      o      o      o      o      sehr deutlich

Der zweite Teil der Onlinebefragung wurde als Hörexperiment konzipiert. Dazu wurden bereits im Laufe der Erhebung zur Drittvariablenkontrolle zwei randomisierte Testgruppen gebildet, um den grundlegenden Anforderungen eines Experiments zu entsprechen.

Beide Testgruppen erhielten die gleichen sechs Hörbeispiele. Diese bestanden aus max. zweiminütigen Solodarbietungen, eingespielt von Studierenden mit Hauptfach Violine. Je zwei Einspielungen stammten von Studierenden der Schulmusik im Bachelor und im Master sowie der Künstlerischen Ausbildung im Master. Das Treatment zielte auf die Testung studienfachbezogener Vorurteile ab (vgl. F 2). Dazu erhielten die Testgruppen unterschiedliche studienfachbezogene Labels zu den Hörbeispielen. Eine Testgruppe erhielt vorweg eine zutreffende Information wie bspw.:

„Auf der folgenden Seite hören Sie eine kurze Aufnahme einer/s Studierenden im **Bachelor Schulmusik** mit Hauptinstrument Violine.

Danach werden Sie gebeten, die Leistung des/der **Bachelorstudent\*in** im Instrumentalspiel einzuschätzen“.

Auf der folgenden Seite konnte die jeweilige Aufnahme genau einmal angehört werden. Die Leistungseinschätzung wurde zur Frage „Wie schätzen Sie die gehörte Leistung des/der Bachelorstudent\*in im Instrumentalspiel nach Ihrem ersten Eindruck ein?“ anhand einer fünfstufigen Skala mit den Extrembeschriftungen *sehr gering* und *sehr hoch* vorgenommen. Außerdem wurde um eine kurze schriftliche Begründung der Einschätzung gebeten.

Die zweite Testgruppe erhielt zur Testung des Vorurteils dagegen eine kontrastierende Angabe zum Studium. Es wurde bspw. ein Masterstudium suggeriert, obwohl das Hörbeispiel von einer Studierenden aus dem Bachelor stammt. Die Hörbeispiele (vgl. Tab. 1) wurden während der Erhebung innerhalb der Testgruppen randomisiert, um Positionseffekte zu vermeiden.

Tabelle 1: Hörbeispiele Bachelor und Master Schulmusik sowie Studiengang Künstlerische Ausbildung (KA) mit echten und falschen Labels für die Testgruppen A und B.

Hörbeispiel	Label Gruppe A	Label Gruppe B
A) Master	Bachelor	Master
B) Bachelor	Master	Bachelor
C) KA	KA	Master
D) Master	Master	KA
E) Bachelor	KA	Bachelor
F) KA	Bachelor	KA

Außerdem unterteilten wir die Stichprobe der befragten Studierenden zur Testung der dritten Forschungsfrage (F3) nach der Expertise im Instrumentalspiel. Die erste Gruppe bestand aus Schulmusikstudierenden versus eine zweite Gruppe zusammengefasst aus Studierenden der Künstlerisch-pädagogischen Ausbildung (KPA) bzw. Künstlerischen Ausbildung (KA). Den Faktor *Expertise* im Instrumentalspiel operationalisieren wir als Unterschied zwischen Studierenden der Schulmusik und Studierenden der KPA (Musikschullehrer\*innen) bzw. KA (angehende Orchestermusiker\*innen). In einer Varianzanalyse mit Messwiederholungen wurde geprüft, ob die beiden Faktoren *Vorurteil* (UV1) und instrumentale *Expertise* (UV2) der Befragten die Bewertungen der sechs Hörbeispiele (AV) beeinflussen.

Die schriftlich gefassten Begründungen zu den Einzelbewertungen wurden zur Beantwortung von Forschungsfrage 4 als Datenbasis für eine qualitative Analyse genutzt. Die Auswertungen wurden mit MAXQDA unter Anwendung der qualitativen Inhaltsanalyse (vgl. Kuckartz & Rädiker, 2022) durchgeführt.

Während die Variable *Bewertung* in der quantitativen Analyse lediglich durch ein einfaches Item operationalisiert wird, erhoffen wir uns durch die qualitative Analyse eine Reihe von differenzierten Einblicken zu den Kriterien der Leistungsbewertung des Instrumentalspiels (vgl. F4). Da in der vorliegenden Stichprobe auch die Angaben von 23 Dozent\*innen vorliegen (s. u.), bot sich flankierend zum vermuteten Einfluss der Expertise (vgl. F3) eine Analyse der qualitativen Daten zu potenziellen Unterschieden in den Bewertungen von Dozent\*innen und Studierenden an.

Zur Evaluation der qualitativen Analysen haben wir abschließend eine KI-gestützte Teilanalyse der qualitativen Daten durchgeführt und diese mit den Analyseergebnissen verglichen (vgl. Rahman et al., 2023). Nach verschiedenen Versuchen erwies sich ChatGPT (Version 4) mit Plugins (PDF, Internetnutzung) für die Methode am besten geeignet.

Zur Teilnahme an unserer Onlinestudie wurden alle 24 deutschen Musikhochschulen angeschrieben. An dem onlinebasierten Experiment nahmen  $N = 313$  Versuchspersonen von 23 Musikhochschulen teil, verteilt auf Studierende der Schulmusik und KPA/KA o.Ä. ( $n = 107$ ), Studierende nur aus Schulmusik-Studiengängen ( $n = 118$ ), Studierende nur aus Künstlerisch-pädagogischer Ausbildung und Künstlerischer Ausbildung ( $n = 65$ ) sowie deren Instrumentaldozent\*innen ( $n = 23$ ). Die aktuellen Studiensemester verteilten sich erwartungsgemäß ohne Auffälligkeiten ( $M = 5.70$ ,  $SD = 3.66$ ).

#### 4. Ergebnisse

*Wie wird die eigene Leistungsentwicklung im Instrumentalspiel von Studierenden der Schulmusik im Studienverlauf eingeschätzt (F1)?*

Die Ergebnisse zur Selbsteinschätzung der instrumentalen Leistungsentwicklung im Schulmusikstudium beziehen sich auf die Teilstichprobe der Studierenden, die nur Schulmusik studieren ( $n = 118$ ) und nicht bereits vorher oder parallel noch KPA/KA studiert haben. Die teilnehmenden Studierenden schätzten die Entwicklung ihrer eigenen instrumentalen Leistung im Studium auf einer 5-stufigen Skala von „gar nicht verbessert“ bis „sehr deutlich verbessert“ zu 71.8% oberhalb des Skalenmittels ein (33.3% auf der höchsten Stufe, 37.8% auf der zweithöchsten Stufe). Mit den restlichen 28.9% der Befragten gab rund ein Drittel lediglich die mittlere oder eine niedrigere Stufe an. Die Ergebnisse zeigen, dass man im Schulmusikstudium nicht von einer Leistungsverbesserung des Instrumentalspiels ausgehen kann.

*Welchen Einfluss haben studienfachbezogene Vorurteile und die Expertise im Instrumentalspiel auf die Bewertung instrumentaler Leistungen (F2 und F3)?*

Der vermutete Einfluss der studienfachbezogenen Vorurteile und der Expertise im Instrumentalspiel auf die Bewertung instrumentaler Leistungen wurde in einer zweifaktoriellen Varianzanalyse mit Messwiederholungen geprüft. Die Stichprobe bestand aus Studierenden, die zum Untersuchungszeitpunkt nur Schulmusik studierten ( $n = 118$ ) oder zusammengefasst nur KPA/KA ( $n = 65$ ). Die Gesamtstichprobe ( $n = 183$ ) ist laut post-hoc-Poweranalyse für eine zweifaktorielle Varianzanalyse mit sechs Messungen geeignet. Die zufallsverteilten Testgruppen für den Faktor *Vorurteil* waren annähernd gleich groß: Gruppe A ( $n = 88$ ) und Gruppe B ( $n = 95$ ). Als Ergebnis ist festzuhalten, dass in der Varianzanalyse mit Messwiederholungen für den Faktor *Vorurteil* kein Unterschied zwischen den Bewertungen der Testgruppen beobachtet wurde ( $F(1,4.95) = 0.29$ ,  $p = 0.59$ , partielles  $\eta^2 = 0.01$ ). Ein angenommener Einfluss von Vorurteilen auf die Bewertung (F2) lässt sich daher nicht bestätigen. Anders verhält es sich in Bezug auf den zweiten Faktor *Expertise*, also potenzielle Unterschiede in der Bewertung der Hörbeispiele in Abhängigkeit der Zugehörigkeit der Probandinnen und Pro-

banden zu den Subgruppen Schulmusikstudierende bzw. der Studierenden aus KPA/KA. Für den Faktor *Expertise* im Instrumentalspiel konnte ein signifikanter Haupteffekt in Bezug auf die Bewertung der Hörbeispiele identifiziert werden ( $F(1,4.95) = 10.06$ ,  $p < 0.01$ , partielles  $\eta^2 = 0.29$ ). Die höchsten Bewertungen (vgl. Abb. 1) erhielten erwartungsgemäß die zwei Hörbeispiele (HB 3 und HB 6), die von Studierenden im Masterstudiengang Künstlerische Ausbildung (KA) eingespielt wurden. Das spricht für die Validität des Verfahrens. Diese Beispiele wurden von Schulmusikstudierenden und Student\*innen der KPA/KA gleichermaßen positiv bewertet. Allerdings gibt es beachtliche Unterschiede in der Bewertung der Hörbeispiele, wenn diese von Studierenden der Schulmusik eingespielt wurden. Schulmusikstudent\*innen geben für die Beispiele aus der Schulmusik deutlich bessere Bewertungen ab, als ihre Kommiliton\*innen aus den Studienfächern KA/KPA.

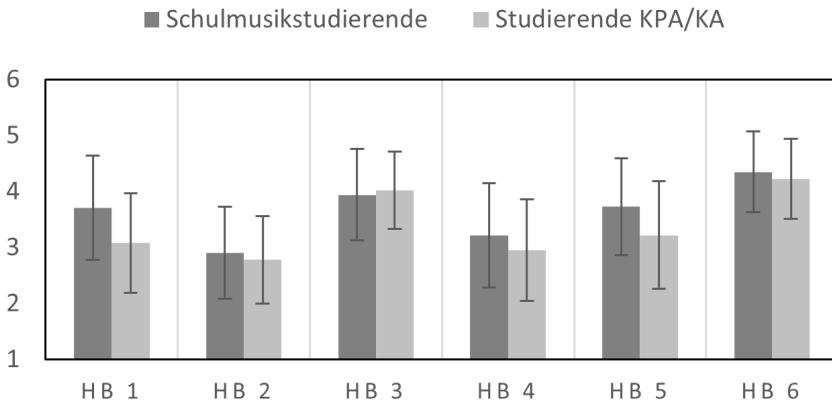


Abbildung 1: Mittelwerte und Standardabweichungen der Bewertungen von sechs instrumentalen Hörbeispielen bei Schulmusikstudierenden verglichen mit Studierenden aus KA/KPA.

Es konnte darüber hinaus kein signifikanter Interaktionseffekt (*Vorurteil* mal *Expertise*) beobachtet werden ( $F(1, 4.95) = 0.32$ ,  $p = 0.90$ , partielles  $\eta^2 < 0.01$ ). Der fehlende Interaktionseffekt zeigt außerdem, dass die hier untersuchten Einschätzungen nicht auf Vorurteilen beruhen. Studierende aus KA/KPA beurteilten die Hörbeispiele nicht schlechter, weil sie aufgrund des Labels vorab von einer instrumentalen Leistung bspw. der Schulmusikstudierenden ausgehen mussten. Das Urteil lässt sich vielmehr auf den Höreindruck zurückführen.



*Welche qualitativen Merkmale spielen bei der Leistungsbewertung des Instrumentalspiels eine Rolle (F4)?*

Die folgende Analyse der qualitativen Daten aller Probandinnen und Probanden zu allen Hörbeispielen zielt auf differenzierte Merkmale ab, die bei der Leistungsbewertung des Instrumentalspiels (hier Hörbeispiele) eine Rolle spielen können. 274 Teilnehmer\*innen gaben insgesamt 1.332 verwendbare textbasierte Begründungen zu ihren Bewertungen ab. Da keine Bewertungskriterien vorgegeben wurden und auch kein allgemeingültiger Kriterienkatalog existiert bzw. überprüft werden sollte, wurde eine induktive Codierung vorgenommen, die zwei Hauptkategorien ergab: a) Musikalische Kriterien und b) Nicht musikalische Kriterien. Die Differenziertheit der Aussagen spiegelt sich in den folgenden Kategorien und Unterkategorien (eingeklammert und der Häufigkeit nach absteigend, Anzahl der Nennungen in eckigen Klammern):

a) Musikalische Kriterien:

- Gestalterische Fähigkeiten (Klang- bzw. Tonerzeugung, Phrasierung, Ausdruck, Vibrato, Musikalität, Dynamik, Artikulation) [263]
- Technische Fähigkeiten (Intonation, Bogentechnik, Fingertechnik) [272]
- Stilistische Korrektheit [40]
- Souveränität in der Ausführung [29]
- Schwierigkeitsgrad des Stückes [41]

b) Nicht musikalische Kriterien:

- Vermeintlicher Studiengang [26]
- Aufnahmequalität des Hörbeispiels [8]
- Ausschnittlänge des Hörbeispiels [8]

Für einen Vergleich der Bewertungen nach Statusgruppen (Dozent\*innen vs. Schulmusiker\*innen) wurden Daten von zufällig ausgewählten 20 Personen pro Gruppe miteinander verglichen. Die Auswertung zeigte, dass insbesondere die Bewertungen der Dozent\*innen negativer bzw. defizitorientierter konnotiert sind (Dozent\*innen: 36 % negative Bewertungen, 26 % positive Bewertungen, 21 % neutrale Bewertungen, 17 % fehlende Bewertungen; Schulmusiker\*innen: 29 % negative Bewertungen, 41 % positive Bewertungen, 22 % neutrale Bewertungen, 8 % fehlende Bewertungen), z. B.:

„Die Klangvorstellung ist eher kratzig, der Bogenstrich nicht besonders geräuscharm. Die Tonqualität ist eher gering“ (587, Pos. 2).

„Unsauber in der Intonation- rhythmisch unsicher“ (698, Pos. 29).

Die Dozent\*innen unterscheiden sich auch in der Wahl der Bewertungskriterien von Studierenden. Die Lehrenden greifen deutlich häufiger auf beobachtbare (operationalisierbare) Kriterien wie intonatorische/rhythmische Ausführung

oder Spieltechnik als auf vergleichsweise mehrdeutige bzw. schwer zu fassende Kriterien wie Musikalität oder Ausdruck zurück. So finden sich in 5,8% der möglichen Bewertungen der Lehrenden schwer zu operationalisierende, aber in 51,7% leichter operationalisierbare Kriterien (Verhältnis 1:8,86). Bei Studierenden liegen diese Häufigkeiten bei 51,7% und 77,5% (Verhältnis 1:1,5).

Für die KI-gestützte Teilanalyse mit Blindversuch wurden aus den qualitativen Daten zwei anonyme Gruppen gebildet: Gruppe A (Dozent\*innen) mit 87 Bewertungseinheiten vs. Gruppe B (Student\*innen) mit 87 Bewertungseinheiten. Die KI erhielt vorab keine Informationen, welche Gruppe aus Dozent\*innen oder Studierenden bestand.

Die KI-gestützten Analyseaufgaben bestanden in der Beschreibung wesentlicher Gruppenunterschiede und mittels Blindversuch in einer begründeten Zuordnung, welche der Gruppen (A oder B) sich aus Studierenden bzw. Dozent\*innen zusammensetzt. Die Zuordnung gelang fehlerfrei und die vorangegangenen Analysen zu den Unterschieden zwischen den Bewertungen der Studierenden und Dozent\*innen wurde mit der KI-gestützten Analyse inhaltlich bestätigt.

Im Gegensatz zu der einfachen Bewertungsskala im Vorurteilsexperiment zeigen die qualitativen Daten, dass die Untersuchungsteilnehmer\*innen bei Gelegenheit auf wesentlich differenziertere Bewertungen zurückgreifen. Skalen mit mehreren Items haben sich in früheren Studien teilweise bewährt (z. B. Harnischmacher & Hofbauer, 2012), teilweise aber auch zu wenig reliablen Ergebnissen geführt, was häufig mit mangelnder Passung von Skala, Bewertungssituation und Stichprobe zusammenhängt (vgl. Hasselhorn & Wolf, 2018). Zur Auswahl geeigneter Items für eine Bewertungsskala für die instrumentale Leistung von Schulmusikstudierenden bieten sich die qualitativen Ergebnisse der vorliegenden Studie an. Ergänzend haben wir daher einen KI-gestützten Itempool generiert. Folgende Kategorien als Grundlagen für Items für eine Bewertungsskala zu Leistungen im Instrumentalspiel (Streicher) wurden über qualitative Inhaltsanalysen und mittels ChatGPT 4 herausgearbeitet, wobei die Items beider Methoden weitgehend übereinstimmen:

1. Intonation
2. Rhythmus
3. Fingertechnik
4. Artikulation
5. Phrasierung
6. Ausdruck
7. Dynamik
8. Musikalische Gestaltung
9. Stilistische Korrektheit
10. Souveränität
11. Tonkontrolle

Die Items sind zunächst nur Vorschläge und müssen im Rahmen einer Skalentwicklung erst empirisch geprüft und evtl. reduziert werden, bevor sie in weiteren Studien eingesetzt werden können.

## 5. Diskussion

Die Ergebnisse legen nahe, dass bei Studierenden der Schulmusik im instrumentalen Hauptfach nicht grundsätzlich von einer Entwicklung der Leistungen vom Bachelor bis zum abschließenden Master auszugehen ist. Offen bleibt, warum fast ein Drittel der befragten Schulmusikstudierenden kaum oder keine Leistungsverbesserung in ihrem Instrumentalspiel beobachten konnten. In Folgestudien wäre zu klären, welche hemmenden Faktoren die Studierenden auf die instrumentale Leistungsentwicklung adressieren. Wir vermuten hier auch Auswirkungen eines Deckeneffekts, wonach ein maximales Leistungsniveau bereits zu Studienbeginn erreicht wurde (vgl. Heyer, 2015). Lehramtsstudierende absolvieren in der Regel nach einer Zeit mit maximalem Übepensum die Eignungsprüfungen. Dieses Pensum können sie dann während des Studiums mit all seinen divergierenden Anforderungen nicht aufrechterhalten. Im Zusammenhang mit der Leistungsmotivation und äußeren Bedingungen (Studienbedingungen, Instrumentalunterricht usw.) könnten auch grundlegende berufliche Zielorientierungen eine Rolle spielen.

Methodisch erscheint neben der Selbstwahrnehmung der Studierenden vor allem auch eine Leistungseinschätzung durch Instrumentaldozent\*innen lohnenswert, um einen noch valideren Blick auf die instrumentale Entwicklung zu erhalten. Zur genaueren Beobachtung der individuellen Leistungsentwicklungen sind dabei für vertiefte und zuverlässige Einblicke Längsschnittstudien notwendig.

Bei der Bewertung der Hörbeispiele konnte sich das studienfachbezogene *Vorurteil* in den Analysen nicht als Einflussgröße durchsetzen. Das ist insofern bemerkenswert, da Vorurteile erfahrungsbasiert sind. Das Ergebnis legt nahe, dass sich Leistungsdifferenzen im Instrumentalspiel zwischen Bachelor- und Masterstudierenden der Schulmusik in den Erfahrungen aus dem Studienalltag gar nicht abbilden. Das zeigt sich auch in der generellen Bewertung der Hörbeispiele. Die Hörbeispiele der Studierenden der Künstlerischen Ausbildung wurden erwartungsgemäß durchweg am besten bewertet. Eine bessere Bewertung der Hörbeispiele aus dem Master im Vergleich zum Bachelor Schulmusik zeigt sich dagegen nicht durchgängig. Diese Beobachtungen lassen sich aufgrund der wenigen Beispiele natürlich nicht verallgemeinern, sprechen aber bei den erwartungsgemäß beobachteten Unterschieden zwischen Studierenden der Schulmusik und der Künstlerischen Ausbildung für die Validität des Verfahrens.

In den qualitativen Bewertungen der Hörbeispiele spielten eine ganze Reihe von Kriterien eine Rolle. Unter methodischen Aspekten ist zu prüfen, ob die ver-

wendete Bewertungsskala (single item) für eine quantitative Analyse zu kurz greift. Möglicherweise lässt sich ein Einfluss studienfachbezogener Vorurteile in einem Folgeexperiment mit einer differenzierten Bewertungsskala besser beobachten, da sich Unterschiede in den Bewertungen eventuell nur in Details, nicht aber in der Gesamtbewertung finden lassen. Ein Itempool für eine mögliche Bewertungsskala wurde durch qualitative und KI-gestützte Analysen entwickelt.

Der Einfluss von *Expertise* auf die Bewertung instrumentaler Leistungen konnte erwartungsgemäß bestätigt werden. Entsprechend der Expertise-Theorie (vgl. Hasselhorn & Knigge, 2018) wurde hier ein klassischer Experten-Novizen-Vergleich vorgenommen. Obwohl die hier operationalisierten Unterschiede als eher klein einzuschätzen sind, konnte festgestellt werden, dass Studierende, die im Rahmen ihres Studiums mehr zeitliche Ressourcen für die Auseinandersetzung mit ihrem Hauptinstrument zur Verfügung haben (KA/KPA) strenger bewerten als Studierende, die ihre zeitlichen Ressourcen mehr verteilen müssen (Schulmusik). Interessant ist dabei vor allem der nicht signifikante Interaktionseffekt. Da sich in der Varianzanalyse kein Haupteffekt für den Faktor *Vorurteil* durchsetzen konnte, wäre ein differenzierterer Vorurteilseffekt z. B. nur für die (vermutlich) leistungsschwächeren Lehramtsstudierenden plausibel gewesen. Stattdessen zeigt sich ausschließlich ein globaler Haupteffekt der Expertise. Entsprechend dieses Befunds zeigen sich dann auch in den qualitativen Daten deutliche Unterschiede in den Bewertungen der Hörbeispiele von Hochschuldozent\*innen im Vergleich zu denen der Studierenden. Die Lehrenden bewerten dabei spezifischer und rekurren eher auf beobachtbare Bewertungskriterien. Im Gegensatz zu Studierenden vermeiden Dozent\*innen Urteile zu übergreifenden Konstrukten wie Musikalität etc., die anhand eines einzigen, kurzen Hörbeispiels als vorschnell bezeichnet werden könnten, sondern bleiben entsprechend ihrer beruflichen Expertise und ihres beruflichen Auftrags bei Kriterien, die im Rahmen von Unterricht auch direkt verbessert werden können. Die KI-gestützte Analyse konnte die Befunde zu den Expertiseunterschieden in einem Blindversuch bestätigen.

## Literatur

- Androutsos, P. & Humphrey, J. T. (2010). Classroom observation ability among preservice music educators in Greece. *International Journal of Music Education*, 28(1), 5–16.
- Behne, K.-E. (1990). Blicken Sie auf die Pianisten?! Zur bildbeeinflussten Beurteilung von Klaviermusik im Fernsehen. *Medienpsychologie*, 2(2), 115–131.
- Bundesverband Musikunterricht e.V. (2023). *Den Musikunterricht sichern. BMU-Position zum Umgang mit dem Musik-Lehrkräftemangel*. [https://www.bmu-musik.de/fileadmin/user\\_upload/Musikunterricht\\_sichern.pdf](https://www.bmu-musik.de/fileadmin/user_upload/Musikunterricht_sichern.pdf)

- Clausen, B. & Schäfer-Lembeck, H.-U. (2017). Passgenauigkeiten. Sondierungen zum Studienerfolg und zum Verbleib von Absolvierenden von Lehramtsstudiengängen an bayerischen Musikhochschulen. *Neue Musikzeitung (nmz)*, 66(6).
- Geuen, H. (2009) »Das kann ich schon!« – Leistungsbewusstsein als Element individueller Lernweggestaltung im Musikunterricht. In H.-U. Schäfer-Lembeck (Hrsg.), *Leistung im Musikunterricht: Beiträge der Münchner Tagung 2008* (S. 55–69). Buch & media.
- Harnischmacher, C., Hasselhorn, J., Schulz-Heidorf, K. & Temming, L. (2021). Individuelle Förderung, Autonomieförderung, Motivation und Musikinteresse in der Familie als Prädiktoren des Kompetenzerlebens und der Zensurenrelevanz im Musikunterricht. *Beiträge empirischer Musikpädagogik*, 12, 1–19.
- Harnischmacher, C. & Hofbauer, V. C. (2012). Musikpädagogisches Handeln und Vorurteil. Eine experimentelle Studie zum Einfluss von Status und Schulform auf die Bewertung von Musikunterricht bei Musiklehramtsstudenten und Schülern. In J. Knigge & A. Niessen (Hrsg.), *Musikpädagogisches Handeln. Begriffe, Erscheinungsformen, politische Dimensionen* (S. 56–69). Die blaue Eule.
- Harnischmacher, C. & Hofbauer, V. C. (2013). War ja klar, dass die nicht unterrichten kann! Eine empirische Folgestudie zum Einfluss von Vorurteilen und Motivation auf die Bewertung von Musikunterricht bei Schülern. *Beiträge empirischer Musikpädagogik*, 4, 1–16.
- Harnischmacher, C. & Knigge, J. (2017). Motivation, Musizierpraxis und Musikinteresse in der Familie als Prädiktoren der Kompetenz „Musik wahrnehmen und kontextualisieren“ und des Kompetenzerlebens im Musikunterricht. *Beiträge empirischer Musikpädagogik*, 8, 1–21.
- Hasselhorn, J. & Knigge, J. (2018). Kompetenz und Expertise. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik* (S. 197–207). Waxmann.
- Hasselhorn, J. & Wolf, A. (2018). Assessment, Bewertung und Musikkritik. In A. C. Lehmann & R. Kopiez (Hrsg.), *Handbuch Musikpsychologie* (S. 389–410). Hogrefe.
- Heyer, R. (2016). *Musiklehramt und Biographie. Rekonstruktion biographischer Orientierungen angehender Musiklehrkräfte*. Springer.
- Hofbauer, V. C., Otradinski, M. & Wehen, A. (2022). *Das künstlerische Hauptfach einfach abwählen? Eine Interviewstudie zur Wahrnehmung der künstlerischen Entwicklung im Hauptinstrument bei Schulmusikstudierenden*. Posterpräsentation auf der Jahrestagung des Arbeitskreises Musikpädagogische Forschung (AMPF), Augsburg. <https://bit.ly/3owmldo>
- Kuckartz, U. & Rädiker, S. (2022). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Beltz.
- Meißner, R. & Schwarz, H. (1999). Ausbildungsgang Schulmusik: Anforderungen, Probleme, Erfahrungen – und ein Spiel mit Worten und Zahlen. In F. Niermann (Hrsg.), *Erlebnis und Erfahrung im Prozess des Musiklernens. (Fest-)Schrift für Christoph Richter* (S. 185–206). Wißner.
- Niessen, A. (2020). Perspektiven auf das Thema Leistung im Musikunterricht. Eine Situationsanalyse auf Basis qualitativer Interviews. *Beiträge empirischer Musikpädagogik*, 11, 1–36. <https://doi.org/10.62563/bem.v2020177>

- Rahman, M., Terano, H. J. R., Rahman, N., Salamzadeh, A. & Rahaman, S. (2023). ChatGPT and Academic Research: A Review and Recommendations Based on Practical Examples. *Journal of Education, Management and Development Studies*, 3(1).
- Richter, C. (2018). Über den Nutzen der künstlerischen Studienfächer im Musiklehramtsstudium. *Diskussion Musikpädagogik*, 77, 10–14.
- Ricken, N. (2014). Adressierung und (Re-)Signifizierung. Anmerkungen zum Zusammenhang von sozialer Herkunft und schulischer Leistung aus praktikentheoretischer Perspektive. In B. Kleiner & N. Rose (Hrsg.), *(Re-)Produktion von Ungleichheiten im Schulalltag: Judith Butlers Konzept der Subjektivierung in der erziehungswissenschaftlichen Forschung* (S. 119–133). Budrich.
- Roberts, B. (1991). Music Teacher Education as Identity Construction. *International Journal of Music Education*, 18, 30–39.
- Switlick, B. & Bullerjahn, C. (1999). Ursachen und Konsequenzen des Abbruchs von Instrumentalunterricht. Eine quantitative und qualitative Umfrage bei Studierenden der Universität Hildesheim. In N. Knolle (Hrsg.), *Musikpädagogik vor neuen Forschungsaufgaben* (S. 167–195). Die blaue Eule.
- Wolf, A. & Kopiez, R. (2014). Do grades reflect the development of excellence in music students? The prognostic validity of entrance exams at universities of music. *Musicae Scientiae*, 18(2), 1–17.

Christian Harnischmacher  
harnischmacher@udk-berlin.de  
<https://orcid.org/0000-0003-2985-7350>

Mascha Otradinski  
mascha-otradinski@gmx.de

Johannes Hasselhorn  
johannes.hasselhorn@fau.de  
<https://orcid.org/0000-0003-3384-224X>

Miriam Knebusch  
miriam.knebusch@fau.de