

*Lukas Bugiel, Julia Bali-Weber,
Malte Sachsse, Katharina Bradler (Hrsg.)*

47. Jahresband des Arbeitskreises Musikpädagogische Forschung

**47th Yearbook of the German Association
for Research in Music Education**

Julia Ehninger

Leistungsbewertung von Gestaltungsaufgaben im Musikunterricht

Perspektiven von Lehrkräften zum Umgang mit einer Herausforderung

*Assessing creative tasks in the music classroom
Teachers' perspectives on navigating a challenge*

Creative tasks such as songwriting, film scoring or creating soundscapes often take place in music classrooms and must be assessed by music teachers. This study examines how teachers navigate this challenge. Seven interviews and two group discussions with music teachers in North Rhine-Westphalia, Germany, were analyzed using the documentary method. Across cases, teachers referred to transparency and fairness as institutional expectations shaping their assessment. Three practices of navigating these expectations emerged: (1) criteria-based assessment, (2) assessment of written tasks and social behavior, and (3) avoiding the assessment of aesthetic quality.

Einleitung

Im Musikunterricht werden regelmäßig Gestaltungsaufgaben bearbeitet, beispielsweise wenn Schüler*innen einen Song schreiben, eine Filmszene vertonen oder eine Klangcollage entwickeln. Solche Aufgaben gelten als besonders kreativ, offen und prozessorientiert. Zugleich steht am Ende eine schulische Anforderung: Es gehört zu einer Kernaufgabe von Lehrkräften, Leistungen zu beurteilen (KMK, 2022). Daher gehen zahlreiche Studien der Frage nach, wie Leistungen – auch im Musikunterricht – möglichst objektiv bewertet werden können. Die Perspektive von Lehrkräften ist dabei bislang jedoch kaum berücksichtigt worden. Hier setzt dieses Forschungsprojekt an: Es untersucht die Herausforderung der Leistungsbewertung von Gestaltungsaufgaben aus der Sicht von Musiklehrkräften.

1. Theoretischer Hintergrund und Forschungsstand

Leistungsbewertung ist in den meisten Schulformen i. d. R. verpflichtend. Zahlreiche empirische Studien belegen eine geringe Reliabilität von Schulnoten – auch in vermeintlich ‚objektiveren‘ Fächern wie bspw. Mathematik (Bohl, 2019, S. 420). Dabei wurde immer wieder gezeigt, dass u.a. subjektive Sichtweisen von Lehrkräften und soziale Bedingungen erheblichen Einfluss auf das Zustandekommen der Note haben (Bohl, 2019, S. 420; Ricken, 2018, S. 51). Gerade Schulleistungsstudien wie *VERA* oder *PISA* haben es sich zur Aufgabe gemacht, Leistungen bzw. Kompetenzen möglichst objektiv zu messen. Auch innerhalb der Musikpädagogik finden sich entsprechende Bestrebungen (im Überblick Hasselhorn & Knigge, 2026). Zugleich gibt es eine anhaltende Debatte über alternative Bewertungsformen, wie z.B. Portfolios oder weitere formative Ansätze (Fautley, 2024; Winter, 2022).

Im Kontext von Gestaltungsaufgaben im Musikunterricht stellen sich zusätzliche Herausforderungen bei der Leistungsbewertung. Hier finden musikalisch-ästhetische Prozesse statt, die mitsamt den entstandenen Produkten bewertet werden. Dabei muss eine Lehrkraft musikalisch-ästhetisch urteilen, wobei solche Urteile in hohem Maße subjektiv sein können (Rolle, 2008). Schulnoten, die ohnehin schon stark von der subjektiven Sichtweise der Lehrperson geprägt sind, könnten dadurch ‚noch subjektiver‘ werden. Innerhalb der Musikpädagogik sind vielfältige Ansätze entwickelt worden, kreative Leistungen zu bewerten (z.B. Deutsch, 2016; Hasselhorn, 2022; Kupers et al., 2018). Jedoch zeigen viele Arbeiten, dass fachlich-inhaltliche Aspekte bei der Leistungsbewertung teilweise nur eine untergeordnete Rolle spielen (z.B. Niessen, 2020 für den Musikunterricht; siehe auch Wagener, 2020 für den Kunstunterricht).

Einem Verständnis von Leistung als individuell erfassbarer Eigenschaften liegt die Annahme zugrunde, dass schulische Leistung beobachtbar und durch geeignete Verfahren objektiv und reliabel messbar ist. Dem stehen Ansätze gegenüber, die schulische Leistung als soziale Konstruktion verstehen (Rabenstein et al., 2015; siehe auch Heberle, 2019, S. 169–179). Rabenstein et al. (2015) etwa gehen davon aus, dass Leistungsbewertung in einem „Konglomerat pädagogischer Praktiken hervorgebracht“ (S. 246) wird. In diesem (praxeologischen) Blick auf Leistungsbewertung ist ausschlaggebend, dass Objektivität von Leistungsbewertung durch solche Praktiken hergestellt wird (Rabenstein et al., 2015, S. 243). So konnte etwa Breidenstein (2023) zeigen, dass es den Beteiligten einer Zeugnisnotenbesprechung durchaus bewusst ist, dass im Kontext schulischer Leistungsbewertung Objektivität nicht erreichbar ist (S. 24). Vielmehr wird eine ‚Objektivität‘ von Schulnoten in sozialen Praktiken erzeugt. Zensuren sind dann ‚objektiv‘, wenn sie in „komplexen sozialen Praktiken objektiviert worden sind“ (Breidenstein, 2023, S. 26).

Leistungsbewertung in der Schule entsteht demnach im Zusammenspiel von normativ geforderter ‚Objektivität‘ sowie sozialen Praktiken der Noten- bzw.

Leistungszuschreibung. Vor diesem Hintergrund geht diese Arbeit nun der Frage nach, wie Lehrkräfte mit der Leistungsbewertung von Gestaltungsaufgaben umgehen.

2. Methode und Fragestellung

Um die Perspektive von Musiklehrkräften zu erheben, wurden sieben leitfaden-gestützte Interviews mit narrativen Anteilen sowie zwei Gruppendiskussionen mit drei bzw. vier Musiklehrkräften aus NRW geführt. Die Gruppendiskussionen dienten hierbei der Erweiterung und komparativen Absicherung der in den Interviews rekonstruierten Orientierungsgehalte. Die teilnehmenden Lehrkräfte waren überwiegend an Gymnasien und Gesamtschulen tätig (Erhebungszeitraum: April 2024 bis Mai 2025) (Tab. 1).

Tabelle 1: Übersicht Stichprobe

Gespräch			Pseudonym	Schulform	Gespräch			Pseudonym	Schulform
Interviews	1		Frau Ahrens (B1)	Gymnasium	Gruppendiskussionen	1		Herr Hess (B8)	Gymnasium
	2		Herr Bahr (B2)	Gymnasium				Herr Ilg (B9)	Gymnasium
	3		Frau Chapo (B3)	Gymnasium				Frau Jütte (B10)	Gymnasium
	4		Frau Dalik (B4)	Förderschule		2		Herr Kuhn (B11)	Gymnasium
	5		Frau Eckert (B5)	Gymnasium				Herr Lucca (B12)	Gesamtschule
	6		Herr Falk (B6)	Gymnasium				Frau Mutz (B13)	Gesamtschule
	7		Frau Götz (B7)	Gymnasium				Herr Noé (B14)	Gesamtschule

Die Gesprächsdaten wurden schließlich mit der Dokumentarischen Methode ausgewertet, da diese „einen Zugang nicht nur zum reflexiven, sondern auch zum handlungsleitenden Wissen der Akteure und damit zur Handlungspraxis“ eröffnet (Bohnsack et al., 2013, S. 9). Die Dokumentarische Methode unterscheidet unter Rückbezug auf die Wissenssoziologie Karl Mannheims zwei Wissens Ebenen, die die Handlungspraxis strukturieren. Auf der einen Seite steht das *kommunikative Wissen*, das auf der Ebene von Alltagstheorien angesiedelt ist und Regeln, institutionalisierte normative Erwartungen sowie Rollenverständnisse umfasst

(u. a. Bohnsack, 2017, S. 103). Dieses Wissen ist Akteur*innen bewusst zugänglich und kann verbal reflektiert werden. Demgegenüber steht das *konjunktive Wissen*, das meist implizit bzw. atheoretisch vorliegt (Bohnsack, 2017, S. 103). Dieses Wissen ist atheoretisch, weil es Personen in ihrer Handlungspraxis anwenden, ohne dass es versprachlicht werden müsste (Nohl, 2017, S. 6). Die beiden Wissens Ebenen können in einem Spannungsverhältnis zueinander stehen, welches das professionalisierte Handeln bestimmt. Dabei wird die Bewältigung dieses Spannungsverhältnisses zwischen „Norm und Praxis oder Norm und Habitus“ (Bohnsack et al., 2024, S. 11) in der Dokumentarischen Methode auch als *Orientierungsrahmen im weiteren Sinne* bezeichnet. Solche Spannungsverhältnisse können z. B. entstehen, wenn Alltagstheorien oder Normen einer Institution in Konflikt mit dem Habitus oder einer Handlungspraxis stehen (Bonnet & Hericks, 2019, S. 106). Für den Kontext Schule bedeutet dies, dass sich das professionalisierte Handeln von Lehrkräften innerhalb dieses Spannungsverhältnisses vollzieht und dass Lehrkräfte Routinen entwickeln, um mit dem Spannungsverhältnis umzugehen (Asbrand & Martens, 2018, S. 23).

Im Anschluss an Bohnsack lassen sich die Normen einer Institution bzw. die „Erwartungen der Lehrperson dessen, was die Institution von ihr erwartet“ (Bonnet & Hericks, 2019, S. 106) als „institutionalisierte normative Erwartungen“ fassen (Bohnsack, 2017, S. 103). Im weiteren Verlauf dieses Beitrags wird *institutionalisierte normative Erwartungen* daher als zentrale analytische Kategorie herangezogen. Zugleich werden – im Anschluss an die Unterscheidung von kommunikativem und konjunktivem Wissen – Orientierungsmuster rekonstruiert, wie Lehrkräfte diese Erwartungen bearbeiten. Diese Orientierungsmuster werden im Folgenden als *Umgangsweisen* bezeichnet.

Bei der Auswertung der Gesprächsdaten wurden die beiden Wissens Ebenen des kommunikativen und konjunktiven Wissens mit den zwei aufeinanderfolgenden Auswertungsschritten *formulierende* und *reflektierende Interpretation* rekonstruiert. Die in diesen beiden Interpretationsschritten rekonstruierten Orientierungen der Akteur*innen wurden schließlich in einer komparativen Analyse über alle Fälle hinweg verglichen.

Vor diesem Hintergrund richtet sich die Analyse erstens auf die Rekonstruktion jener *institutionalisierten normativen Erwartungen*, die Lehrkräfte im Zusammenhang mit der Leistungsbewertung von Gestaltungsaufgaben benennen (Forschungsfrage I) und zweitens auf den Umgang der Lehrkräfte mit diesen Erwartungen (Forschungsfrage II):

- I. Welche institutionalisierten normativen Erwartungen prägen das professionalisierte Handeln von Musiklehrkräften im Kontext der Leistungsbewertung von Gestaltungsaufgaben?
- II. Mit welchen Umgangsweisen begegnen Lehrkräfte diesen Erwartungen in ihrer Bewertungspraxis?

3. Ergebnisse

Zunächst stelle ich in Kapitel 3.1 einen exemplarischen Gesprächsausschnitt aus den Daten sowie dessen Interpretation vor. Der Gesprächsausschnitt dient dazu, das Vorgehen der Interpretation zu veranschaulichen, auf das sich die anschließende Ergebnisdarstellung ab Kapitel 3.2 stützt. In dieser Ergebnisdarstellung werden zunächst die institutionalisierten normativen Erwartungen rekonstruiert, die das professionalisierte Handeln von Musiklehrkräften im Kontext der Leistungsbewertung von Gestaltungsaufgaben prägen (Kap. 3.2, siehe auch Forschungsfrage I). In Kapitel 3.3 werden schließlich drei Umgangsweisen rekonstruiert, die Lehrkräfte nutzen, um diesen institutionalisierten normativen Erwartungen gerecht zu werden (siehe auch Forschungsfrage II).

3.1 Gesprächsausschnitt

Der folgende Gesprächsausschnitt stammt aus einem Interview mit Frau Ahrens, die seit etwa zwei Jahren als Musiklehrerin an einem Gymnasium arbeitet. In dieser Passage beschreibt sie die letzte Situation in ihrem Unterricht, in der sie Gestaltungsaufgaben bewertete. In einer siebten Klasse hatten Schüler*innen zuvor Werbevideos zu einem bestehenden oder selbst ausgesuchten Werbeartikel gedreht.

1 Dann war natürlich auch die Frage wie bewerte ich das denn? gebe ich
 2 da jetzt irgendwie Noten dafür? un=und ich hab mich dann aber erstmal
 3 so ein bisschen dagegen entschieden da die tatsächlich mit einer Note
 4 zu bewerten, sondern eher so mit so einer Tendenz, ich habe dann so
 5 ein Kompetenz-, ähm (2) Kompetenzraster? irgendwie so gemacht, wo ich
 6 dann geguckt hab, ham=ham=ham se selbst komponierte Musik da drin
 7 gehabt, ham se (.) ähm (2) irgendein passendes, ähm (2) Produkt
 8 irgendwie? ham, halt mit solchen Sachen halt geguckt und das so
 9 bewertet und halt eher so anhand von von so Smileys dann halt irgendwie
 10 dann halt das so bewertet, wo ich für mich dann halt schon weiß, okay,
 11 das geht eher so in in einen sehr guten Bereich, //mhm// und das war
 12 (.) auch gu::t, das war n biss- ja so ja so zwischendrin zwischen gut
 13 und befriedigend und so, (.) aber ich habe jetzt (.) ein bisschen
 14 Scheu gehabt, da tatsächlich irgendwie den Leuten hier so eine du hast
 15 jetzt eine eins minus [haut auf den Tisch] und du hast ne zwei plus [haut auf
 16 den Tisch] //mhm// oder so, weil ich da echt so ein bisschen (2) weil im
 17 Prinzip waren se alle gut. un=un=und dann natürlich fangen die dann
 18 auch an zu vergleichen //mhm// und ich habe dann halt auch so so kleine
 19 Zettelchen dann halt auch gemacht und (2) kommt natürlich dann immer
 20 auch [spricht schneller und ahmt ihre Schüler:innen nach] "ja, aber warum, aber wir
 21 hatten doch hier und dann hatten wir doch da auch, warum also das
 22 nicht da angekreuzt?" und so un=un (2) wo ich dann auch wieder gemerkt
 23 hab, das ist total schwierig da ähm (2) tatsächlich (.) also für mich
 24 (.) das irgendwie so gerecht zu machen, //mhm// dass irgendwie alle
 25 damit //mhm// einverstanden sind, und das dann auch so transparent wie
 26 möglich zu machen. //mhm// ist es dann okay, d=d=da so viel Transparenz
 27 dann herzustellen? dass die dann sofort anfangen zu vergleichen?
 28 wa=was (.) welches Feld ist bei euch angekreuzt und welches bei uns
 29 und so? (.) ja (.) schwierig.

Abbildung 1: Gesprächsausschnitt Frau Ahrens

Auf *kommunikativer Ebene* zeigt sich, dass Frau Ahrens mit dem „Kompetenzraster“ (Z. 5) ein Bewertungsinstrument nutzt und damit eine kriteriengeleitete, für Schüler*innen nachvollziehbare Bewertung realisieren möchte. Die genannten Kriterien (z.B. „selbst komponierte Musik“, „passendes [...] Produkt“, Z. 6–7) ermöglichen es ihr, einen Sachbezug herzustellen. Mithilfe der „Zettelchen“ (Z. 19) wird die Bewertung der Videoclips für die Schüler*innen nachvollziehbar und sie schafft so Transparenz. In diesem Vorgehen dokumentiert sich eine Orientierung an den *institutionalisierten normativen Erwartungen* von *Fairness* (sie möchte es „gerecht“ machen, Z. 24), *Nachvollziehbarkeit* (alle sollen „damit einverstanden“ sein, Z. 25) und *Transparenz* (Z. 26–27), wobei sie diese als Norm zugleich auch infrage stellt, da sie dazu führt, dass sich Schüler*innen vergleichen.

Auf *konjunktiver Ebene* verweist der erste Satz des Gesprächsausschnitts („Dann war natürlich auch die Frage wie bewerte ich das denn?“, Z. 1) auf die Selbstverständlichkeit, mit der die Leistungsbewertung zur Unterrichtspraxis dazugehört. Gleichzeitig dokumentiert sich in der gesamten Passage eine latente Unsicherheit, was sich u.a. an der wiederholten Verwendung von „irgendwie“ (Z. 2, 5, 8, 9, 14, 24) festmachen lässt. Zunächst zeigt sich ein Widerstreben von Frau Ahrens, die Videoclips mit einer Ziffernote zu bewerten. Sie entscheidet sich „ein bisschen dagegen“ (Z. 3) und bewertet lediglich „mit so einer Tendenz“ (Z. 4). Trotz dieses Widerstrebens dokumentiert sich in der gesamten Passage

die Orientierung, Produkte in besser und schlechter einzuteilen. Dies zeigt sich u. a. an besagter „Tendenz“ sowie auch an der Verwendung von „Smileys“ (Z. 9). Die Kriterien, die sie beispielhaft anführt (Z. 6–7), können mit ‚ja‘ oder ‚nein‘ beantwortet werden, sind also dichotom und direkt beobachtbar. Obwohl sich Frau Ahrens ursprünglich gegen eine Bewertung in Form von Ziffernnoten entschieden hat, kann sie durch dieses Vorgehen die Leistung in einen Notenbereich überführen (Z. 11–13). Schließlich leitet sie performativ einen Gegenhorizont zu ihrer ursprünglich intendierten Praxis ein: Sie haut in einer karikierenden Geste gleich zweimal auf den Tisch, wodurch sie die Notengebung richterlich als autoritär und für Schüler*innen nicht nachvollziehbar markiert (Z. 15–16). Diese körperliche Geste deutet als performativer Akt auf konjunktive Wissensbestände hin, und so wird nonverbal eine Distanz zu einer autoritär-intransparenten Notenvergabe sichtbar. Damit referenziert sie eine nicht nachvollziehbare Bewertungspraktik, von der sie sich eigentlich abgrenzen möchte. Mit der Aussage „im Prinzip waren sie alle gut“ (Z. 16–17) markiert sie schließlich die Nicht-Vergleichbarkeit der Produkte und relativiert eine Einteilung in bessere und schlechtere Produkte.

Das professionalisierte Handeln von Frau Ahrens orientiert sich an den schulischen Normen von Transparenz und Fairness, und diesen Normen wird sie gerecht, indem sie kriteriengeleitet bewertet. Dennoch zeigt sich Frau Ahrens' Widerstreben, die Werbevideos der Schüler*innen zu bewerten. In der gesamten Passage dokumentieren sich somit Spannungen sowohl innerhalb als auch zwischen den Wissens Ebenen. Schulische Bewertung ist nicht nur als institutionalisierte normative Erwartung auf der kommunikativen Ebene präsent, sondern dokumentiert sich zugleich auf konjunktiver Ebene als habitualisierte Verständlichkeit („natürlich“, Z. 1). Auf kommunikativer Ebene wird deutlich, dass Videoclips nicht vergleichbar sind („im Prinzip waren sie alle gut“, Z. 16–17) und dass diese kriteriengeleitete Bewertung, die eine Vergleichbarkeit zwischen den Produkten herstellt, nicht aufgeht, da sich die Qualität der Videoclips nicht in Kriterien überführen lässt. Damit kollidieren hier Wissensbestände um die Nicht-Vergleichbarkeit von ästhetischen Produkten mit der gleichzeitigen Pflicht zur Vergleichbarkeit durch Kriterien und Noten.

3.2 Institutionalisierte normative Erwartungen

Fallübergreifend ließ sich im Kontext der Leistungsbewertung von Gestaltungsaufgaben eine Orientierung an den institutionalisierten normativen Erwartungen von *Transparenz* und *Fairness* rekonstruieren (siehe Forschungsfrage I). Leistungsbewertung soll transparent und nachvollziehbar sein, und diese Erwartung prägt das professionalisierte Handeln der befragten Musiklehrkräfte maßgeblich. Eine solche transparente Bewertungspraxis wird in den Gesprächen immer wieder einer nicht nachvollziehbaren Bewertungspraxis gegenübergestellt. Frau

Ahrens etwa wollte unbedingt vermeiden, dass die Schüler*innen nur eine Ziffernote zu ihren Videoclips erhalten. In einer Gruppendiskussion wird die Bewertungspraxis bei *Jugend musiziert* als negativer Gegenhorizont zu einer transparenten Leistungsbewertung herangezogen. Laut Frau Jütte sei *Jugend musiziert*

ja der Klassiker davon, dass man es eigentlich nicht versteht, warum man jetzt den dritten Preis bekommen hat oder die Weiterleitung oder so, weil es eben keine (.) keine Kriterien gibt. ich bin kriterienlos aufgewachsen, (.) und immer nur mit einem Ergebnis von=@frontiert@ worden (.) konfrontiert worden.

Es zeigt sich fallübergreifend, dass diese Norm der Transparenz auch von den Schüler*innen eingefordert wird. Zudem wird in allen Gesprächen die Norm der Fairness verhandelt und es wird v. a. darüber diskutiert, dass es schwierig ist, dieser Norm gerecht zu werden. Als Gründe hierfür wird u.a. die unterschiedliche musikalische Vorbildung von Schüler*innen genannt, weiterhin die Schwierigkeit, Leistungen in Gruppenarbeiten zu bewerten, sowie die hohe Anzahl von Schüler*innen im Klassenverband.

3.3 Umgangsweisen mit institutionalisierten normativen Erwartungen

Aus allen Gesprächsdaten konnten schließlich drei Umgangsweisen rekonstruiert werden, mit denen Musiklehrkräfte den institutionalisierten normativen Erwartungen von *Transparenz* und *Fairness* bei der Leistungsbewertung von Gestaltungsaufgaben begegnen (Forschungsfrage II): (a) *Kriteriengeleitete Bewertung*, (b) *Bewertung von Ersatz- und Zusatzleistungen* und (c) *Vermeidung der Bewertung musikalisch-ästhetischer Qualität*. Diese drei Umgangsweisen wurden fallübergreifend in der komparativen Analyse rekonstruiert und stellen typische Bearbeitungsformen von Herausforderungen im Zusammenhang mit institutionellen normativen Erwartungen und alltäglicher Unterrichtspraxis dar. Sie umfassen dabei Wissensbestände der kommunikativen und der konjunktiven Ebene.

(a) *Kriteriengeleitete Bewertung*

Alle Musiklehrpersonen orientieren sich bei der Leistungsbewertung von Gestaltungsaufgaben an *kriteriengeleiteter Bewertung*. Dabei erscheint kriteriengeleitete Bewertung in den Gruppendiskussionen als *der Weg*, Gestaltungsaufgaben zu bewerten. In den Gesprächen wird immer wieder auf „die“ Kriterien verwiesen, ohne dass dabei erläutert wird, was damit konkret gemeint ist. Dies deutet daraufhin, dass das Wissen um „die“ Kriterien Teil des beruflichen Erfahrungswissens der Lehrpersonen ist, das zumindest teilweise implizit vorliegt. Zugleich erscheint es als Alltagstheorie: Durch Kriterien wird Notengebung objektiver, fairer und transparenter. Zwar besteht Einigkeit über das grundlegende Prinzip einer kriteriengeleiteten Bewertung, allerdings nicht über die inhaltliche Festle-

gung der Kriterien: Deren konkrete Ausgestaltung bleibt vielfach offen, und „die“ Kriterien bleiben inhaltlich weitestgehend unbestimmt. Wenn Kriterien explizit benannt werden, sind dies häufig eher musikalisch-technische Kriterien, die direkt beobachtbar sind und bei denen es ein „richtig“ oder „falsch“ geben kann (wie z. B. in Abb. 1, Z. 6–7).

Die Orientierung an solchen Kriterien scheint aus Sicht der Lehrpersonen die institutionalisierten normativen Erwartungen von *Transparenz* und *Fairness* zu erfüllen, weil die Bewertung anhand von Kriterien die Gestaltungsprodukte vergleichbar macht. Gleichzeitig dokumentiert sich in den Gesprächen immer wieder eine Unsicherheit darüber, ob „die“ Kriterien die musikalisch-ästhetische Qualität eines Produktes abbilden. Laut Herrn Ilg kann ein Gestaltungsprodukt auch „cool“ sein, wenn es nicht „die“ Kriterien erfüllt (siehe auch [c]). Kriteriengeleitete Bewertung erscheint damit als eine pragmatische Umgangsweise, um institutionalisierten normativen Erwartungen gerecht zu werden. Zusammengefasst erscheinen „die“ Kriterien als allgegenwärtige Legitimationsressource auf einer alltagstheoretischen Ebene: Sie lösen den Anspruch von Transparenz und Vergleichbarkeit ein, bleiben inhaltlich jedoch häufig unbestimmt.

(b) Bewertung von Ersatz- und Zusatzleistungen

Statt des eigentlichen Produkts von Gestaltungsaufgaben bewerten die Lehrkräfte häufig *Ersatz- oder Zusatzleistungen*, wie Reflexionen, Protokolle, Storyboards oder die Dokumentation des Produkts. So verweist etwa Herr Ilg darauf, dass er bei Gestaltungsaufgaben „auch schriftlich was verlange“ und er bei Songwritingprojekten „eigentlich nicht“ bewerte, „wie gut die Songs sind“. In allen Gesprächen dokumentiert sich die Alltagstheorie, wonach *schriftliche Leistungen* transparenter und fairer seien. Entsprechend werden solche schriftlichen Leistungen häufig als Bewertungsgrundlage von Gestaltungsaufgaben herangezogen. Schriftliche Leistungen lassen sich abgeben oder nicht; wenn sie nicht abgegeben werden, „ist es halt schlecht“ (Herr Ilg).

Zudem zeigt sich in allen Gesprächen, dass das *Arbeits- und Sozialverhalten* der Schüler*innen Grundlage von Leistungsbewertung ist. In vielen Gesprächen wird ein deutlicher Gegensatz aufgemacht zwischen engagierten und nicht engagierten Schüler*innen, wie hier bei Frau Eckert: „die Dödel die nichts machen und die Leute, die (.) die sich total engagieren“. Das Arbeits- und Sozialverhalten von Schüler*innen wird in Alltagstheorien zu einem beobachtbaren Kriterium stilisiert, das bewertbar ist und somit auch der institutionalisierten normativen Erwartung von Transparenz gerecht wird. Gleichzeitig wird immer wieder die Problematik dieser Praxis reflektiert, weil sie mit *Fairness* in Konflikt stehen kann. In vielen Gesprächen wird etwa darüber sinniert, ob es fair ist, jemandem eine bessere Note zu geben, nur weil diese Person mehr „Spaß“ hat.

Um die institutionalisierten normativen Erwartungen von *Transparenz* und *Fairness* einzulösen, wird die Bewertung vom ästhetischen Produkt weg auf diejenigen Anteile verschoben, die abgabefähig, sprachlich begründbar oder

beobachtbar sind. Damit entsteht eine ‚Verschiebung von Leistung‘ weg vom eigentlichen Gegenstand der Gestaltungsaufgabe hin zu kognitiven bzw. prozess-bezogenen und sozialen Leistungen.

(c) Vermeidung der Bewertung musikalisch-ästhetischer Qualität

In vielen Gesprächspassagen zeigt sich schließlich eine Vermeidung, die musikalisch-ästhetische Qualität der Gestaltungsprodukte zu bewerten. Dies wird jedoch nur von den wenigsten Lehrkräften explizit benannt. Gleichwohl teilen die Musiklehrkräfte in den Gruppendiskussionen das Wissen darüber, dass sich die musikalisch-ästhetische Qualität von Gestaltungsprodukten nicht immer in Kriterien überführen lässt. Ein Produkt kann „cool“ sein, auch wenn es nicht die Kriterien erfüllt (Herr Ilg, siehe auch [a]). Diese Wissensbestände zeigen sich in verschiedenen Bewertungspraktiken, die sich in den Gesprächen dokumentieren. Einzelne Lehrkräfte bewerten ausschließlich die Ersatz- oder Zusatzleistungen (siehe [b]), andere vergeben Extrapunkte für kreative Ideen. Herr Falk begründet dies damit, dass dies im Abitur auch so gemacht werde, und sagt hierzu: „Ich kann doch nicht bewerten, ob ähm, ein Mensch jetzt (.) zufällig kreativer ist als der andere.“ Zugleich scheint die musikalisch-ästhetische Qualität ohnehin immer wieder eine untergeordnete Bedeutung zu spielen, da Schüler*innen Kriterien, so wie sie auf dem freien Markt existieren, schlichtweg nicht erfüllen können. Wenn dann doch die musikalisch-ästhetische Qualität beurteilt wird, wird mit den institutionalisierten normativen Erwartungen gebrochen. Hierzu Herr Lucca:

Ich muss sagen, ich stimme euch voll zu. das Allerwichtigste natürlich sind die Kriterien und so und auf der anderen Seite irgendwie, (.) ich ertappe mich dabei auch, (.) dass (.) also keine Ahnung jetzt mal so als Beispiel? (.) jemand hat sich jetzt nicht so ganz? krass an diese ähm Kriterien gehalten, aber das Ergebnis ist einfach so geil, dass man sich denkt, [aaaaah, als Geräusch] (.) ob man, (.) also ich bin da also ein bisschen durchlässiger dann, und dann denke ich mir so okay, ähm, ich weiß nicht, dass das so ein bisschen [anonymisiert] das klingende Ergebnis zählt.

Die Beteiligten wissen, dass eine Bewertung der musikalisch-ästhetischen Qualität in Spannung zu den institutionalisierten normativen Erwartungen von Transparenz und Fairness steht, und umgehen diese Spannung, indem sie ästhetische Qualität entweder ausklammern oder nur zusätzlich (als Bonus) berücksichtigen. Wenn – wie im Beispiel von Herrn Lucca – die ästhetische Qualität dennoch ausschlaggebend wird („das Ergebnis ist einfach so geil“), wird bewusst von der schulischen Erwartung einer kriteriengeleiteten Bewertung abgewichen.

4. Zusammenfassung und Diskussion

Die Auswertung der Gesprächsdaten zeigt, dass das professionalisierte Handeln von Musiklehrkräften bei der Leistungsbewertung von Gestaltungsaufgaben von institutionalisierten normativen Erwartungen geprägt ist. Fallübergreifend ließ sich eine Orientierung an den institutionalisierten normativen Erwartungen von *Transparenz* und *Fairness* rekonstruieren (vgl. Kap. 3.2). Obwohl sich die Lehrkräfte immer wieder kritisch mit Leistungsbewertung auseinandersetzen, wird Notengebung nicht grundsätzlich in Frage gestellt (wie im Gespräch von Frau Ahrens, Abb. 1).

Im komparativen Vergleich konnten drei Umgangsweisen mit institutionalisierten normativen Erwartungen rekonstruiert werden (vgl. Kap. 3.3): *(a) Kriteriengeleitete Bewertung*, *(b) Bewertung von Ersatz- und Zusatzleistungen* und die *(c) Vermeidung der Bewertung musikalisch-ästhetischer Qualität*. Diese drei Umgangsweisen sind nicht als voneinander unabhängige Handlungsoptionen zu verstehen, sondern als unterschiedliche Bearbeitungsformen wiederkehrender, teils konfligierender Wissensbestände: Auf kommunikativer Ebene teilen die Lehrkräfte die institutionalisierten normativen Erwartungen von Transparenz und Fairness sowie die Alltagstheorie, dass kriteriengeleitete Bewertung diesen Erwartungen gerecht wird. Zugleich zeigt sich ebenfalls auf kommunikativer Ebene, dass sich musikalisch-ästhetische Qualität nicht ohne Weiteres in Kriterien und Noten überführen lässt – auch wenn ein Produkt „geil“ (siehe oben) sein kann. Außerdem dokumentiert sich auf konjunktiver Ebene eine habitualisierte Bewertungspraxis des Einteilens in bessere und schlechtere Leistungen sowie das Nicht-Bewerten von Leistungen.

Teil des konjunktiven Erfahrungsraumes der Lehrkräfte ist die geteilte Ungewissheit, wie institutionalisierten normativen Erwartungen begegnet werden kann. Diese Ungewissheit wird bearbeitet, indem bekannte schulische Bewertungslogiken reproduziert werden. Kriteriengeleitete Bewertung scheint dabei auf kommunikativer Ebene *der Weg*, institutionalisierten normativen Erwartungen gerecht zu werden. Allerdings bleibt das, was sich hinter „den“ Kriterien verbirgt, häufig implizit oder es werden eher beobachtbare, teils dichotome Kriterien angeführt. Dies lässt auf geteilte Wissensbestände darüber schließen, dass musikalisch-ästhetische Qualität mit den Normen schulischer Leistungsbewertung nicht vereinbar ist. Zudem lässt sich in dieser Arbeit ein Phänomen beobachten, das in kompositionspädagogischen Arbeiten beschrieben wurde: Pädagog*innen halten sich häufig in schulischen Kontexten bei der Bewertung musikalisch-ästhetischer Qualität zurück. Julia Weber (2021, S. 167–180) spricht in diesem Zusammenhang auch von einem Spannungsfeld von Kunst und Pädagogik, in dem pädagogisches Handeln stattfindet: Kunst wird mit Freiheit assoziiert, und als Konsequenz kommt es zu einer Zurückhaltung, sich als Pädagog*in in künstlerische Prozesse von Lernenden einzumischen.

Einschränkend muss festgehalten werden, dass die beobachteten Orientierungsmuster möglicherweise schulformspezifisch für Gymnasium und Gesamtschule sind. Frau Dalik, die früher an einem Gymnasium arbeitete und nun an einer Förderschule tätig ist, macht deutliche Unterschiede zwischen den Schulformen auf. Die Ergebnisse sind auch darüber hinaus kontextgebunden: Die Perspektive von Schüler*innen und Unterrichtsbeobachtungen fehlen. Dies ist bedeutsam vor dem Hintergrund, dass viele Arbeiten zeigen konnten, wie ein Verständnis von Leistung im schulischen Kontext von Lehrkräften und Schüler*innen gemeinsam hervorgebracht wird (u.a. Breidenstein, 2012; Rabenstein et al., 2015; Wagoner, 2020).

Die Befunde bestätigen Annahmen innerhalb der Professionsforschung, die davon ausgehen, dass der Umgang mit Ungewissheit zur Kernaufgabe professionellen (pädagogischen) Handelns gehört (u.a. Helsper, 2008; Paseka et al., 2018) und dass diese Ungewissheit im Handeln mit der Doppelstruktur aus kommunikativem und konjunktivem Wissen einhergeht (Bohnsack et al., 2022, S. 17). Aus dieser Doppelstruktur heraus ergibt sich auch die Frage, was eigentlich als Leistung im Kontext von Schule bzw. Gestaltungsaufgaben verstanden wird. Das Verständnis, das in den Gesprächen an den Tag gelegt wird, ist ein Verständnis von Anstrengung. Leistungsbewertung scheint so viel weniger mit den individuellen musikalischen Kompetenzen der Schüler*innen zu tun zu haben, als dass sie sozial konstruiert wird (Ricken, 2018) und durch eine ‚Scheinobjektivität‘ aufrechterhalten wird.

Literatur

- Asbrand, B. & Martens, M. (2018). *Dokumentarische Unterrichtsforschung*. Springer.
- Bohl, T. (2019). Leistungsbewertung, Notengebung und Alternativen zur Notengebung. In E. Kiel, B. Herzig, U. Maier & U. Sandfuchs (Hrsg.), *Handbuch Unterrichten an all-gemeinbildenden Schulen* (S. 414–425). Klinkhardt.
- Bohnsack, R. (2017). *Praxeologische Wissenssoziologie*. Barbara Budrich.
- Bohnsack, R., Bonnet, A. & Hericks, U. (2022). Praxeologisch-wissenssoziologische Professionsforschung. Rahmung und Erträge einer feldübergreifenden Perspektive. In R. Bohnsack, A. Bonnet & U. Hericks (Hrsg.), *Praxeologisch-wissenssoziologische Professionsforschung. Perspektiven aus Früh- und Schulpädagogik, Fachdidaktik und Sozialer Arbeit* (S. 13–30). Klinkhardt.
- Bohnsack, R., Nentwig-Gesemann, I. & Nohl, A.-M. (Hrsg.). (2013). *Die dokumentarische Methode und ihre Forschungspraxis. Grundlagen qualitativer Sozialforschung* (3. Aufl.). Springer.
- Bohnsack, R., Sturm, T. & Wagoner, B. (2024). Konstituierende Rahmung und professionelle Praxis. Pädagogische Organisationen und darüber hinaus. Ein erster Überblick. In R. Bohnsack, T. Sturm & B. Wagoner (Hrsg.), *Konstituierende Rahmung und professionelle Praxis. Pädagogische Organisationen und darüber hinaus* (S. 9–16). Barbara Budrich. <https://doi.org/10.25656/01:28804>

- Bonnet, A. & Hericks, U. (2019). Professionalität und Professionalisierung als biographische Bearbeitung der Spannungen zwischen Norm und Habitus. In R.-T. Kramer & H. Pallesen (Hrsg.), *Lehrerhabitus. Theoretische und empirische Beiträge zu einer Praxeologie des Lehrerberufs* (S. 101–123). Klinkhardt.
- Breidenstein, G. (2012). *Zeugnisnotenbesprechung. Zur Analyse der Praxis schulischer Leistungsbewertung*. Barbara Budrich.
- Breidenstein, G. (2023). Zensuren: Risiken und Nebenwirkungen. In C. Führer, M. Magirus, T. Bohl, B.-S. Grewe, W. Polleichtner & F. Ulfat (Hrsg.), *Relativität und Normativität von Beurteilungen. Fächerübergreifende und fachspezifische Analysen* (S. 23–30). Tübingen Library Publishing. <http://dx.doi.org/10.15496/publikation-85924>
- Deutsch, D. (2016). Authentic Assessment in Music Composition: Feedback That Facilitates Creativity. *Music Educators Journal*, 102(3), 53–59. <https://doi.org/10.1177/0027432115621608>
- Fautley, M. (2024). Assessment of Composing in the Lower Secondary School in England. In K. Devaney, M. Fautley, J. Grow & A. Ziegenmeyer (Hrsg.), *The Routledge Companion to Teaching Music Composition in Schools. International Perspectives* (S. 452–465). Routledge.
- Hasselhorn, J. (2022). Kreativität im Musikunterricht in Zeiten von Kompetenzorientierung. Entwicklung eines Modells kreativer Kompetenz in Musik. In M. Göllner, J. Knigge, A. Niessen & V. Weidner (Hrsg.), *43. Jahresband des Arbeitskreises Musikpädagogische Forschung* (S. 159–174). Waxmann.
- Hasselhorn, J. & Knigge, J. (2026). Kompetenz und Expertise. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Hrsg.), *Handbuch Musikpädagogik* (2. Aufl., S. 385–398). Waxmann.
- Heberle, K. (2019). *Zur Konstruktion von Leistungsdifferenz im Rahmen musikpädagogischer Unterrichtspraxis. Eine Videostudie zum instrumentalen Gruppenunterricht in der Grundschule*. Waxmann.
- Helsper, W. (2008). Ungewissheit und pädagogische Professionalität. In Bielefelder Arbeitsgruppe 8 (Hrsg.), *Soziale Arbeit in Gesellschaft* (S. 162–168). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90960-8_18
- KMK. (2022). *Standards für die Lehrerbildung: Bildungswissenschaften* (Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 07.10.2022). https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- Kupers, E., Van Dijk, M. & Lehmann-Wermser, A. (2018). Creativity in the Here and Now: A Generic, Micro-Developmental Measure of Creativity. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.02095>
- 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.v2020>
- Nohl, A.-M. (2017). *Interview und Dokumentarische Methode. Anleitungen für die Forschungspraxis* (5. Aufl.). Springer.
- Paseka, A., Keller-Schneider, M. & Combe, A. (Hrsg.). (2018). *Ungewissheit als Herausforderung für pädagogisches Handeln*. Springer. <https://doi.org/10.1007/978-3-658-17102-5>
- Rabenstein, K., Idel, T.-S. & Ricken, N. (2015). Zur Verschiebung von Leistung im individualisierten Unterricht. Empirische und theoretische Befunde zur schulischen Leitdif-

- ferenz. In J. Budde, N. Blasse, A. Bossen & G. Reißler (Hrsg.), *Heterogenitätsforschung. Empirische und theoretische Perspektiven* (S. 241–259). Beltz Juventa.
- Ricken, N. (2018). Konstruktionen der ‚Leistung‘. Zur (Subjektivierungs-)Logik eines Konzepts. In S. Reh & N. Ricken (Hrsg.), *Leistung als Paradigma: Zur Entstehung und Transformation eines pädagogischen Konzepts* (S. 43–60). Springer VS. <https://doi.org/10.1007/978-3-658-15799-9>
- Rolle, C. (2008). Argumentationsfähigkeit: eine zentrale Dimension musikalischer Kompetenz? In H.-U. Schäfer-Lembeck (Hrsg.), *Leistung im Musikunterricht. Beiträge der Münchner Tagung 2008* (S. 70–100). Allitera.
- Wagener, B. (2020). *Leistung, Differenz und Inklusion. Eine rekonstruktive Analyse professionalisierter Unterrichtspraxis*. Springer.
- Weber, J. (2021). *Stimmigkeit und Resonanz. Zum Zusammenhang zwischen Überzeugungen von Komponist*innen und ihrem kompositionspädagogischen Handeln*. Waxmann.
- Winter, F. (2022). *Leistungsbewertung. Eine neue Lernkultur braucht einen anderen Umgang mit den Schülerleistungen* (8. Aufl.). Schneider.

Julia Ehninger
 Bergische Universität Wuppertal
 Gaußstraße 20
 42119 Wuppertal
 ehninger@uni-wuppertal.de
<https://orcid.org/0000-0003-0554-7630>