

Lingua musica? Zur Erfassung musiksprachlicher Kompetenzen Jugendlicher in textbasierten Testinstrumenten

Michael Ahlers & Andreas Seifert

Zusammenfassung

Der Artikel beschreibt die Entwicklung eines Instruments zur Erfassung von (Teil-)Kompetenzen Jugendlicher in der siebten Klasse in den Bereichen Musik-Fachsprache sowie des Sprach- und Leseverstehens. Durch die Verwendung von authentischen Texten in eigens konstruierten C-Tests wurden Daten aus zwei Stichproben ($N_{\text{gesamt}} = 299$) gewonnen und analysiert. Die Güte der Subtests ist dabei gut bis sehr gut. Ein dreidimensionales Modell bildet die Zusammenhänge zwischen den notwendigen allgemeinen sprachlichen und den (musik-)fachsprachlichen Kompetenzen am besten ab. Dabei kann durch den Einsatz von Regressionsanalysen gezeigt werden, dass es einen musikspezifischen Bereich von Fachsprache geben kann. Varianzanalysen bestätigen, dass die Schulform und Muttersprachlichkeit der Probanden wichtige Einflussgrößen dieser erfassten Teilkompetenz sind. Abschließend werden die Optionen eines Einsatzes des Instruments mit Bezug zu bereits vorliegenden Instrumenten und Konstrukten sowie die weitere Validierung des Instrumentes und des zugrunde liegenden, Rasch-skalierten Modells diskutiert. Es wird reklamiert, dass in der Konstruktion von musikbezogenen Testinstrumenten und kompetenzorientierten Aufgaben zu selten auf empirisch gesichertes Wissen um Fachsprachlichkeit Wert gelegt wird.

Abstract

The article illustrates the development of an instrument for the measurement of music specific language and competences. It focuses on students of the seventh grade in Germany. C-tests were constructed by selection and use of authentic schoolbook texts. The reliability of all sub scales is good to very good. Scales were scaled by use of the Rasch-model. A three-dimensional model is able to represent the correlations and coherences between general literacy, and special musical literacy, or music specific language competence. The use of regression analysis proved this special language to exist. Analysis of variance showed the effects of the test person's mother language and the type of school they attended. The instrument is able to augment knowledge on musical competences, and provides an economic way to collect data on this. Furthermore it can be correlat-

ed to other constructs and testing tools. The authors argue that implementing knowledge on the structure and use of music specific language could result in the construction of better surveys, or educational material.

1 Einleitung

1.1 Verortung des Vorhabens

Die folgenden Ausführungen bewegen sich in einem wissenschaftlich gesehen interdisziplinären, bildungs- wie kulturpolitisch relevanten Feld, da einerseits Kompetenz- und Leistungsmessungen innerhalb pädagogischer und psychologischer Kontexte einbezogen und andererseits hierin auch Diskurse um Heterogenität und Diversität auf den Ebenen *Bildung* sowie *Gesellschaft* tangiert werden. Hierbei ist bemerkenswert, dass den sprachlichen Fähigkeiten und Fertigkeiten von Probanden bisher in musikpädagogischen wie musikpsychologischen Testinstrumenten vergleichsweise wenig Aufmerksamkeit gewidmet wurde. Generell erscheint die *sprachliche Schwierigkeit* von Aufgaben oder Instruktionen in den genannten Disziplinen noch ein wenig besetztes Feld zu sein. Es werden nachfolgend sprachliche Schwierigkeiten in der Bewältigung von Textaufgaben abgebildet, die sich in die Bereiche allgemeiner sprachlicher sowie bildungs- und fachsprachlicher Kompetenzen differenzieren (Becker-Mrotzek et al., 2013; Rautenberg, 2012) und in Bezug zu weiteren Variablen setzen lassen. Hierdurch sollen einige „blinde Flecken“ über die Kenntnisse und Fähigkeiten Jugendlicher im Alter zwischen 12 und 15 Jahren und die Zielgruppen-Adäquatheit der aktuell zugelassenen bzw. genutzten Vermittlungstexte und Testaufgaben im schulischen Musikunterricht grundlegend bearbeitet werden.

Aber auch für die angewandte Musikpsychologie können diese Ergebnisse im Rahmen der Konstruktion von Testinstrumenten fruchtbar sein: Denn differenziertes Wissen über die Struktur der musikalischen Fachsprache und ihr Verhältnis zu allgemeinen Sprach- und Lesefähigkeiten kann und sollte bei der Testkonstruktion qualitätssteigernd wirken.

Bekannt, aber nachfolgend nicht weiter referenziert, sind an dieser Stelle diejenigen, auch im deutschen Sprachraum vorfindlichen, umfangreichen Arbeiten zur Verbindung sowie Verortung von „Musik und Sprache“, in denen sowohl philosophische wie literatur-, sprach-, kommunikations- oder neurowissenschaftliche Diskurse stattfinden.

1.2 Stand der Forschung

Neben der Erfassung von Musikalität über die gängigen Instrumente, etwa von Carl E. Seashore, Arnold Bentley oder Edwin E. Gordon (vgl. von Georgi, 2010) sowie der Intelligenzdiagnostik ist allgemein die Kompetenzmessung im Jahr 2014 auch im Umfeld der Musikwissenschaft und Musikpädagogik ein äußerst lebendiger Bereich: Reinhard Kopiez und Anna Wolf (2014) stellten erste Er-

gebnisse ihres Vorhabens zur Entwicklung eines empirisch geprüften Kompetenzmodells und modellbasierter Aufgaben für Musiktheorie und Gehörbildung an Hochschulen vor, Johannes Hasselhorn und Andreas C. Lehmann (2014) optimieren und theoretisieren *KOPRA-M*¹ weiter und auf der Jahrestagung des *Arbeitskreises Musikpädagogische Forschung (AMPF)* widmeten sich mehrere Vorträge unter anderem der Erfassung ästhetischer Argumentationskompetenz sowie den möglichen Bezügen von und zwischen Konstrukten wie Motivation, Begabung oder musikalischen Kompetenzen (zusammengefasst bei Niessen & Knigge, im Druck).

Die Legitimation dieser vielfältigen Zugänge und fachwissenschaftlichen wie fachdidaktischen Projekte ist dabei bildungspolitisch erklärbar: Im Zuge der internationalen Leistungsstudien wie „Programme for International Student Assessment (PISA)“ (Deutsches PISA-Konsortium, 2000) oder „Trends in International Mathematics and Science Study (TIMSS)“ wurde klar, dass deutsche Schüler zwar Routine- und Reproduktionsaufgaben gut lösen können, sie aber schwache Leistungen bspw. im Bereich des Leseverstehens und des Transfers zeigen. Durch die Entwicklung nationaler Bildungsstandards (Klieme et al., 2007) sowie die Beschreibung und Erfassung fachspezifischer Kompetenzmodelle sollte dieser Situation begegnet werden.

Kompetenzen werden dabei nach Franz Weinert (2001) verstanden als

„die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ (Weinert, 2001, S. 27)

Diese starke Gewichtung von Kompetenzen erklärt sich daraus, dass Kompetenzen kognitionspsychologisch als notwendige Bedingung für flexibles und adaptives Handeln in komplexen Anforderungssituationen angesehen werden und damit genau demjenigen Leistungsverhalten zu Grunde liegen, das die deutschen Schüler in internationalen Leistungsstudien eher nicht abrufen konnten (vgl. Schneider & Stern, 2010, zit. nach Gogoll, 2014, S. 2).

Während diese Haltung für viele Fächer inzwischen als akzeptiert angesehen werden kann, tat und tut sich die Forschung mit Bezug auf Musik vergleichsweise schwer, in ihrer grundlegenden Behandlung des Themas auf einen Konsens in der Beschreibung/Formulierung und Testung eines übergreifenden Kompetenzmodells zu gelangen oder generell den Begriff der Kompetenz zu adaptieren (Knigge, 2014). Lediglich das *KoMus*-Projekt, welches an der Universität Bremen koordiniert wurde, lieferte ein empirisch validiertes Teil-Kompetenzmodell. Innerhalb der Definition dieses Teil-Kompetenzmodells für den Musikunterricht und dessen Validierung (Jordan, 2014; Jordan et al., 2012; Knigge, 2011; Niessen et al., 2008) sowie den bisher vorliegenden, empirisch nicht fundierten Teilkompetenz-Modellen sensu Jank (2005) oder internationalen Standards werden zwar auch sprachliche Kompetenzen zumindest ansatzweise thematisiert,

1 Vgl. <http://www.hfm-wuerzburg.de/kopra-m.html> [02.03.2015].

beispielsweise auf den drei Niveaus „basal“, „mittel“ und „differenziert“ (Niesen et al., 2008, S. 20). Im Unterschied zu der nachfolgend beschriebenen Studie fokussieren die Autorinnen und Autoren von *KoMus* jedoch die Verbalisierung von Wahrnehmungen und schließen dabei auf implizite und explizite Wissensbestände der Probanden (Jordan et al., 2012, S. 505). Die Sprachstruktur der Aufgaben selbst wird hinsichtlich ihrer fachwissenschaftlichen wie fachsprachlichen Schwierigkeiten nicht theoretisch fundiert analysiert, sondern nur durch Experten bewertet. Das Leseverstehen (von Aufgaben) und dessen späterer Transfer werden unserer Auffassung nach nicht ausreichend berücksichtigt.

Um jedoch diese Wissenslücken um die *realen* Sprach- und vor allem Lese- und Verständnisschwierigkeiten sowie das fachsprachliche Vokabular von Schülern erfassen und in Beziehung zu weiteren Faktoren wie der Schulform, den Erst- und Zweitsprachen, dem Geschlecht oder fachspezifischen Fertigkeiten und Fähigkeiten setzen zu können, wurde ein eigenes Testinstrument entworfen und im Rahmen der Pilotierungsstudie erprobt.

Mittelfristig kann nach Auffassung der Autoren aus pädagogischer Sicht auf einer in dieser Weise fundierten Daten- und Wissensbasis mit der angemessenen Konstruktion oder Revision kompetenzorientierter Lern- und Leistungsaufgaben begonnen werden. In der weiteren Beforschung des Konstrukts musikalischer, fachsprachlicher Kompetenzen ist es so nachfolgend möglich, dieses Instrument an Intelligenz-Diagnostika (z. B. den *BEFKI*) oder weitere Instrumente zur musikalischen Teilkompetenz-Messung anzubinden oder es diesen voranzustellen.

2 Methode

Die Studie lässt sich in zwei Erhebungen zu unterschiedlichen Messzeitpunkten einteilen. Die Erhebungen sind als Pilotierungen des Modells zu verstehen und unterscheiden sich hinsichtlich der Konstruktion der Testhefte, nicht aber in der Art der verwendeten Test-Texte. Vor den Erhebungen wurden die Tests durch Angehörige des Instituts für Kunst, Musik und ihre Vermittlung der Leuphana Universität in Lüneburg mit Studierenden pre-getestet und inhaltlich validiert.

In den Erhebungen wurden sogenannte C-Tests genutzt, mithilfe derer allgemeine linguistische und semantische Fähigkeiten über die Reduktion von Redundanzen erfasst werden. C-Tests sind vielfach erprobte und hoch zuverlässige Instrumente (Baur, Grotjahn & Spettmann, 2006; Grotjahn 1994, 2002; Harsch & Hartig, 2010), welche vor allem innerhalb der Fremdsprachen-Didaktik, in Tests zu „Deutsch als Fremdsprache“ oder im „Test of English as a Foreign Language (*TOEFL*)“-Test eingesetzt werden.

Benötigt werden zur Konstruktion jeweils authentische Texte mit einer vorgegebenen Mindestlänge, welche innerhalb der „kanonischen“ oder konservativen Konstruktion über die Tilgung von 25 Wortendungen (jedes zweite Wort) systematisch beschädigt werden. Hierbei geben die Längen der Lücken bewusst keinerlei Aufschluss über die Anzahl der fehlenden Buchstaben.

An Abbildung 1 lassen sich die Dimensionen der getesteten Kompetenzbereiche gut illustrieren: Um diesen Text rekonstruieren zu können, sind sowohl

Die Geige

Die Geige oder Violine ist ein Instrument, das vielseitig verwendet wird. Die Gei~~ge~~ gehört – w~~erden~~ die Brat~~sche~~, das Cel~~lo~~ und d~~er~~ Kontrabass – z~~ur~~ Familie d~~er~~ Streichinstrumente. Der T~~on~~ entsteht, w~~enn~~ die Sai~~ten~~ zum Schw~~innen~~ gebracht wer~~den~~. Dies ka~~nn~~ durch Stre~~ken~~ der Sai~~ten~~ mit ei~~ner~~ Bogen o~~der~~ durch Anzu~~g~~ erreicht wer~~den~~. Unterschiedliche Tonhö~~hen~~ werden du~~rch~~ die L~~ä~~, die Di~~cke~~ und Span~~nung~~ der Sai~~ten~~ erzeugt. Mit Hilfe der Wirbel stimmt man die Saiten.

Abb. 1:

Exemplarischer C-Text zur Geige mit Antwortverhalten eines Probanden

allgemein-sprachliche Fähigkeiten notwendig aus den Bereichen Lexik, Syntax oder Kohäsion (Beispiele: „der“, „zur“ usw.), sowie auch fachsprachliches Wissen. Dieses kann einerseits ebenfalls lexikalische Wissensbereiche tangieren (bspw. „Saiten“, „Tonhöhen“), aber auch aus den Bereichen Semantik oder Morphosyntax stammen (bspw. wie Saiten gespielt werden können, welche Parameter die Tonhöhe beeinflussen können usw.). Die konservative Tilgung der Texte erfasst ein höheres Maß an allgemeinsprachlichen Fähigkeiten als bspw. eine reine Tilgung des Wortbeginns von Fachbegriffen. In Prätests wurde deutlich, dass die zuletzt genannte Konstruktion zu einem erheblichen Anstieg der Schwierigkeit und zu massiven Zeitproblemen während der Testung führt. Daher wurde im weiteren Verlauf auf die konservative Konstruktion der Subtests zurück gewechselt.

Alle 25 Lücken der so beschädigten Texte fungieren im weiteren Verlauf als Items. Diese wurden im Rahmen dieser Studie nur als gelöst/nicht gelöst kodiert, Schreibweisen der Probanden wurden (weitgehend) vernachlässigt. Eine nicht bearbeitete Lücke wurde als nicht gelöst kodiert. Eine differenziertere Kodierung ist möglich, hier jedoch nicht angewendet worden. Insgesamt stehen den Probanden für die Texte jeweils fünf Minuten Bearbeitungszeit zur Verfügung. Wie sich ebenfalls an Abbildung 1 ablesen lässt, hat dieser Proband ausreichend Zeit gehabt (letzte Lücke wurde bearbeitet), konnte aber – wahrscheinlich aufgrund fehlender fachsprachlicher Fähigkeiten – die drei Lücken davor nicht lösen.

Im Vorfeld der Konstruktion des musikspezifischen Erhebungsinstruments wurde zur Herstellung der curricularen Validität eine Analyse des niedersächsischen Kerncurriculums sowie des Bildungsplans Hamburg durchgeführt, um thematische Schwerpunkte der fünften und sechsten Klassen im Fach Musik zu extrahieren. Es wurde dabei klar, dass sich länderspezifische Bildungspläne hinsichtlich einiger Formulierungen unterscheiden, die grundlegenden inhaltlichen Dimensionen aber weitgehend übereinstimmend definiert werden und nicht gravierend voneinander abweichen. Die siebte Klasse wurde als Zielgruppe der Erhebung ausgewählt, da davon auszugehen ist, dass in allen Schulformen mindestens in der fünften und sechsten Klasse noch Musik unterrichtet wird. Später

ist regelmäßig davon auszugehen, dass dieses Fach entweder halbjährlich wechselt mit dem Fach Kunst oder gänzlich entfällt.

Anschließend wurden Musik-Schulbücher nach verwertbaren Texten durchsucht, um hierdurch inhaltliche Validität zu gewährleisten und letztlich mit authentischen Textbausteinen arbeiten zu können. Als Filterkriterien für die Auswahl der Bücher fungierten hierbei deren Erscheinungsjahr, die Zulassung für die Unterstufe (Klassen 5 und 6) sowie die Textlängen der Fachtexte selber (Näheres hierzu siehe Ahlers & Seifert, in Druck).

Es konnten so insgesamt 12 Texte extrahiert werden, die den genannten Kriterien entsprachen und in denen die in beiden Bundesländern innerhalb der in den Curricula der fünften und sechsten Klasse beschriebenen inhaltlichen Kategorien auftauchten (siehe Tab. 1).

Tab. 1:
Kategorien und Anzahl der extrahierten Texte

Kategorie	Name des Textes	Kurzform (in SPSS)
Musikal. Parameter	1. Metrum und Takt	(A_Takt/A_Takt)
	2. Der Takt	(B_Takt/B_Takt)
Klangphänomene/ Instrumente	3. Die Rohrblattinstrumente	(Blatt/Rohrblattinstr)
	4. Die Klaviermechanik	(Piano/Klaviermechanik)
	5. Die Geige	(Viol/Violine)
Stimmphysiologie	6. Die Stimme	(Vox/Stimme)
Hörphysiologie	7. Das Gehör	(Ohr/Ohr)
Musikal. Formen	8. Formen in der Musik	(A_Form/A_Formen), nur in erster Erhebung
	9. Gestalt und Form	(B_Form/B_Formen), nur in zweiter Erhebung
Kulturelle Dimensionen	1. Was ist ein musikalisches Programm?	(A_Prog/A_Programm)
	11. Dem Programm auf der Spur	(B_Prog/B_Programm), nur in zweiter Erhebung
Grundlagen der Akustik	12. Wie entstehen Klänge und Töne?	(Klang/Klang)

Mit diesen präparierten Texten wurden für die erste Erhebung insgesamt drei Testhefte erstellt, da die Probanden in der zur Verfügung stehenden Zeit nicht alle Items hätten bearbeiten können. Alle Hefte enthielten dabei die Abfrage soziodemografischer Angaben, der Mutter-, Zweit- sowie ggf. weiteren Sprachlichkeit sowie die Abfrage der letzten Benotung im Fach Musik. In den Testheften waren jeweils vier Fachtexte des Bereichs Musik und ein – allen Testheften

gemeinsamer – Kontrolltext aus der Deutsch-Didaktik vorhanden. Somit ergaben sich insgesamt 325 Items.

Für die zweite Erhebung wurde das Instrument in ein Multi-Matrix-Design (Balanced-Incomplete-Block-Design, vgl. von Davier et al., 2006) überführt, um die IRT-Skalierung besser zu gewährleisten (in der ersten Erhebung sind lediglich 3 Testhefte mit jeweils vier Musikaufgaben und einer Deutschaufgabe genutzt worden, sodass keine Ankeritems vorhanden waren).

In der finalen Version der Testhefte wurden 12 musikspezifische sowie der Kontroll-/Anker-Text aus der Deutsch-Didaktik in einer Matrix rotiert, wodurch sich dann über die Aufteilung von fünf Texten pro Heft insgesamt 33 Testhefte ergaben. Das Design ist in Tabelle 2 dargestellt.

Tab. 2:
Balanced Incomplete Block Design der Studie

Testlet	Aufgabe 1	Aufgabe 2	Aufgabe 3	Aufgabe 4	Aufgabe 5
1	A_Takt (1)	B_Takt (2)	Blatt (3)	Piano (4)	Deutsch (13)
2	B_Takt (2)	Blatt (3)	Piano (4)	Viol (5)	Deutsch (13)
3	Blatt (3)	Piano (4)	Viol (5)	Vox (6)	Deutsch (13)
...
12	A_Takt (1)	B_Takt (2)	Deutsch (13)	Blatt (3)	Piano (4)
...
23	Deutsch (13)	A_Takt (1)	B_Takt (2)	Blatt (3)	Piano (4)
...

3 Ergebnisse

3.1 Beschreibung der Stichproben

An der ersten Erhebung nahmen insgesamt 198 Schüler teil (49 % weiblich, 51 % männlich), deren durchschnittliches Alter 13,2 Jahre betrug. Wenngleich es in Norddeutschland nunmehr auch „Oberschulen“ gibt, wurde im Rahmen der Studien zwischen Realschul- (54,5 %) und Hauptschulklassen (31,3 %) differenziert und als weitere Schulform noch die integrierte Gesamtschule (14,1 %) hinzugenommen. 71,7 % der Probanden (Pbn) gaben an, Deutsch als Erstsprache zu sprechen, gefolgt wurde diese Angabe von Russisch (6,1 %) und Arabisch (4,5 %). Unter den immerhin 11,6 % der angegebenen „anderen Erstsprachen“ (offenes Item), rangierten Kurdisch, Albanisch, Englisch und Kroatisch auf den vorderen Plätzen. Unter den Zweitsprachen steht Deutsch ebenfalls auf dem ersten Platz (29,8 %), vor Russisch (2 %) und Polnisch (1,5 %).

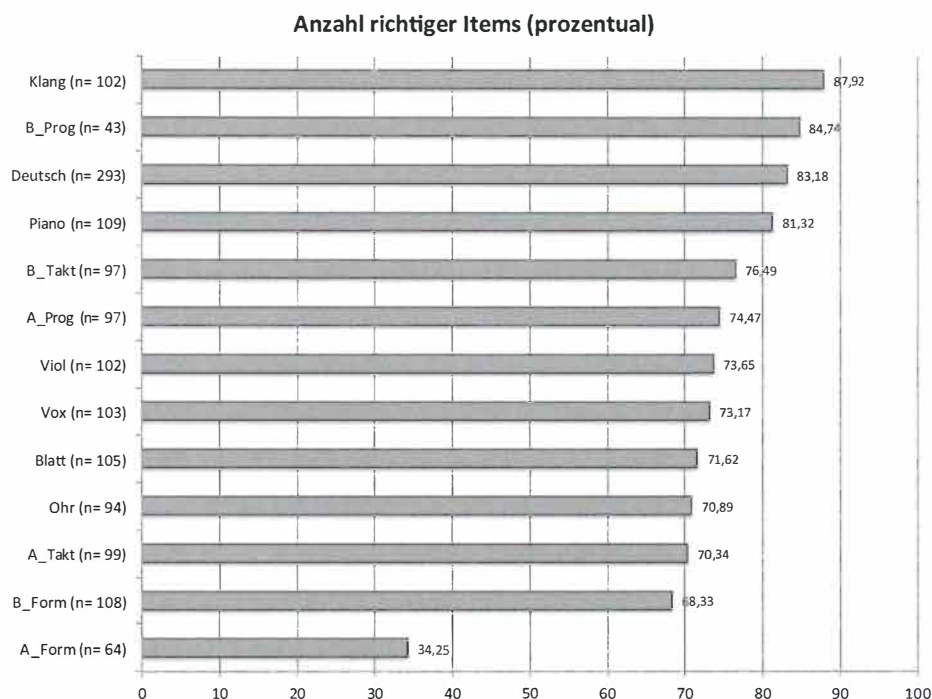


Abb. 2:
Prozentuale Lösungshäufigkeiten der Items pro Skala

Die zweite Erhebung bestand ausschließlich aus gymnasialen Teilnehmern, was sich aus forschungspraktischen Gründen erklären lässt. 93 % der insgesamt 101 Schüler gaben an, Deutsch als ihre Erstsprache zu nutzen, nur 3 % der Pbn gaben an, mit russischer Sprache aufgewachsen zu sein und bei den fehlenden vier Prozent waren dies Türkisch, Montenegrinisch oder Englisch. 28,4 % der Stichprobe verfügen privat über eine Zweitsprache. Hier finden sich am häufigsten Russisch und Polnisch (je 13,2 %), gefolgt von Spanisch und Italienisch. Das durchschnittliche Alter der Stichprobe beträgt 12,89 Jahre. Sie besteht zu zwei Dritteln aus Jungen (67,3 %).

3.2 Deskriptive Daten

Nach einem allgemeinen, deskriptiven Blick auf die Daten überrascht zunächst die scheinbar geringe Schwierigkeit der meisten Skalen hinsichtlich der prozentualen Lösungswahrscheinlichkeit innerhalb der Gesamtstichprobe (siehe Abb. 2). Einzig die Skalen zur Kategorie „musikalische Formen“ weichen hiervon ab. Dies wird auf Seite 235 noch ausführlicher diskutiert.

3.3 Skalierung und Auswertungen

Die Daten beider Teilstudien (N=299) bilden im weiteren Verlauf der Auswertung die Grundlage für eine Rasch-Skalierung (durchgeführt mit der Software *Conquest*, vgl. Wu, Adams & Wilson, 1997). Es wurden zunächst drei unterschiedliche Modelle (eindimensional vs. zweidimensional vs. dreidimensional) gegeneinander getestet. Das erste Modell basiert auf allen 325 Items, also 12 × 25 Items der Aufgaben mit musikalischem Inhalt und eine Aufgabe mit 25 Items „Deutsch“. Im zweiten Modell basiert die erste Dimension auf 300 Items mit musikalischem Inhalt (12 Inhaltsgebiete à 25 Lücken), die zweite Dimension besteht aus 25 Lücken des Deutsch-Texts. Im dritten Modell wird die erste Dimension noch einmal unterteilt: Items, die mit allgemeinen sprachlichen Kenntnissen zu lösen sind (N=184) sowie Items, für deren Lösung explizit musikalisches Wissen notwendig ist (N=116). Die musikalischen Items (N=300) lagen, wie oben beschrieben, durch das Rotationsdesign jeweils nicht allen getesteten Schülern vor, der Deutschtest war in allen Testheften vorhanden.

Die globalen Gütemaße aller Dimensionen aller Modelle weisen auf einen hinreichenden Fit hin (siehe Tab. 3).

Da die globalen Gütemaße der Skalen der drei Modelle eine gute bis sehr gute Zuverlässigkeit andeuten, wurde mithilfe der finalen Devianz und Freiheitsgrade beider Modelle ein χ^2 -Differenztest durchgeführt, um entscheiden zu können, welches Modell Grundlage weiterer Berechnungen sein soll. Zudem wurde als informationstheoretisches Maß der CAIC-Index² hinzugezogen. Die Daten belegen eine bessere Anpassung des zwei- bzw. dreidimensionalen Modells gegenüber dem eindimensionalen Modell (Tab. 4). Während allerdings auf der Grundlage des CAIC das zweidimensionale Modell die beste Anpassung aufweist, zeigt sich beim χ^2 -Differenztest eine Überlegenheit des dreidimensionalen Modells. Den folgenden Berechnungen wird das dreidimensionale Modell zugrunde gelegt.

Tab. 3:
Globale Gütemaße der Skalen beider Modelle

	Modell 1 (1-Dim)	Modell 2 (2-Dim)		Modell 3 (3-Dim)		
Dimension	General-Faktor	Musik	Deutsch	Musik Fach	Musik Allgemein	Deutsch
Varianz	2.375	2.483	2.659	2.872	2.295	2.900
EAP/PV-Reliabilität	.939	.931	.854	.929	.929	.834

Anmerkungen: Die Varianz sollte mindestens 1,0 betragen. EAP/PV= Expected A-Priori/Plausible Values – Die EAP/PV-Reliabilität sollte mindestens .70 betragen.

2 Consistent Akaike Information Criterion (vgl. Rost, 1996, S. 328ff.)

Tab. 4:
 χ^2 -Differenztest der Modelle

	Modell 1 (1-Dim)	Modell 2 (2-Dim)	Modell 3 (3-Dim)
CAIC	26758.8628	26638.03135	26651.12552
Devianz	24621.42130	24480.48852	24466.78092
Parameter	319	322	326
Differenz Devianz	140.93278		13.7076
Differenz Parameter	3		4
<i>p</i>	< .01		< .01

Tab. 5:
Zusammenhänge zwischen den Skalen

	MusikFach	MusikAllgemein	Deutsch
MusikFach	1	.881***	.734***
MusikAllgemein		1	.748***
Deutsch			1

Anmerkung: *** = signifikant auf .001-Fehlerniveau

Im Folgenden wurden WLE-Scores³ für die drei Dimensionen berechnet, die als abhängige bzw. unabhängige Variablen für weitere Analysen herangezogen werden. Die Zusammenhänge zwischen den drei Scores sind in Tabelle 5 abgetragen.

Erwartungsgemäß korrelieren die Subdimensionen des Musiktests signifikant höher untereinander als mit dem Deutschtest ($Z=5.39$, $p<.001$ resp. $Z=5.01$, $p<.001$). Die Zusammenhänge der beiden Subdimensionen mit dem Deutschtest unterscheiden sich hingegen nicht signifikant ($Z=0.38$, $p>.05$).

Schulform

Alle drei Dimensionen zeigen signifikante Mittelwertunterschiede nach Schulform (siehe Tab. 6). Alle Effekte sind als groß zu bezeichnen (vgl. Bortz & Döring, 2006, S. 606). Der größte Effekt zeigt sich in der Dimension „Musik-Fach“.

3 WLE = Weighted Least Square Estimates. Die Ermittlung der WLE-Scores basiert in der IRT-Skalierung auf Punktescores, die Personen in einem Test erreicht haben. Mit Bezug auf diese Punktescores wird die Lösungswahrscheinlichkeit (Personenparameter) errechnet, mit der eine Person ein bestimmtes Item lösen kann. Die jeweiligen Lösungswahrscheinlichkeiten werden logarithmiert, sodass sie Intervallskalenniveau aufweisen. Hohe WLE-Scores stehen für eine hohe Personenleistung.

Tab. 6:
Varianzanalysen (UV: Schulform)

	DF_{Schulform}	DF_{Fehler}	F	p	η^2
MusikFach	3	289	113;714	< .001	.541
MusikAllgemein	3	289	97;401	< .001	.503
Deutsch	3	289	65;197	< .001	.404

Tab. 7:
Deskriptive Statistiken der Skalen nach Schulform

	Musik Fach		Musik Allgemein		Deutsch		
Schulform	M	SD	M	SD	M	SD	N
Hauptschule	5.760	1.300	6.287	1.249	5.577	1.440	62
Realschule	6.976	1.011	7.396	0.954	6.712	1.266	107
Integrierte Gesamtschule	7.783	1.295	7.938	0.996	6.961	1.104	28
Gymnasium	9.130	1.237	9.270	1.233	8.389	1.221	96

Die Unterschiede gestalten sich in allen Dimensionen erwartungskonform, also Hauptschule < Realschule < Integrierte Gesamtschule < Gymnasium (vgl. Tab. 7).

Alter

Die Variable „Alter“ korreliert *negativ* mit allen drei Dimensionen ($r = -.178$, $p < .01$ resp. $r = -.161$, $p < .001$ resp. $r = -.196$, $p < .001$). Dieser kontraintuitive Zusammenhang („je jünger, desto besser“) lässt sich durch die Abhängigkeit der Variable „Alter“ (in dieser Stichprobe) von der Schulform erklären ($F(3;295) = 7.241$, $p < .001$, $\eta^2 = .069$). Wird die Schulform statistisch kontrolliert, ergibt sich ein nichtsignifikanter Effekt des Alters auf alle Dimensionen.

Geschlecht

Die Ergebnisse des deutschen PISA-Konsortiums (2000) und diejenigen in den „Deutsch Englisch Schülerleistungen International (DESI“-Studien (Nold & Willenberg, 2007) sowie neuerer Arbeiten mit Hinblick auf Mutter- und Fremdsprachlichkeit bspw. von Mashkovskaya (2013) ließen erwarten, dass Mädchen signifikant besser hinsichtlich ihrer Sprachkompetenzen sowie fachsprachlicher Fähigkeiten abschneiden als Jungen. Der Einfluss des Geschlechts wird jedoch in keiner der drei Dimensionen innerhalb der Stichproben signifikant ($F(1;290) = .078$, $p > .05$ resp. $F(1;290) = .80$, $p > .05$ resp. $F(1;290) = .487$, $p > .05$).

Tab. 8:
Varianzanalysen (UV: Muttersprache)

	DF_{Muttersprache}	DF_{Fehler}	<i>F</i>	<i>p</i>	η^2
MusikFach	1	291	48;401	< .001	.143
MusikAllgemein	1	291	54;331	< .001	.157
Deutsch	1	291	36;666	< .001	.112

Tab. 9:
Deskriptive Statistiken der beiden Skalen nach Muttersprache

	MusikFach		MusikAllgemein		Deutsch		
Sprache	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>N</i>
Deutsch	7.840	1.637	8.151	1.472	7.329	1.489	231
Andere	6.241	1.487	6.619	1.382	5.985	1.772	62

Muttersprache

Der Mittelwertunterschied bezüglich der Muttersprache ist in der vorliegenden Stichprobe in allen Dimensionen signifikant (siehe Tab. 8). Die Effektgrößen sind in etwa vergleichbar, allerdings zeigt sich, dass die Größe der Effekte im Verhältnis zu denen der Schulform eher als gering einzuschätzen ist.

In allen drei Skalen zeigen Deutsch-Muttersprachler bessere Leistungen als Nichtmuttersprachler. Tabelle 9 zeigt hierzu die Mittelwerte und Standardabweichungen.

Spezifität der Skala „MusikFach“

Um den Anteil spezifischer Varianz der Skala „MusikFach“ zu ermitteln – um also sagen zu können, inwiefern spezifisches musikalisches Wissen notwendig ist, welches über reine sprachliche Kompetenz hinausgeht, um die Items, die als „Fachsprache“ identifiziert wurden, erfolgreich zu beantworten – wird eine Regressionsanalyse mit den unabhängigen Variablen „MusikAllgemein“ und „Deutsch“ und der abhängigen Variablen „MusikFach“ berechnet. Das Gesamtmodell weist einen hohen Anteil an aufgeklärter Varianz auf ($R^2 = .789$). Jedoch sind 21,1 % der Varianz *nicht* aufgeklärt, was für eine gewisse Spezifität des Konstrukts spricht. Ein Blick auf die standardisierten Regressionsgewichte der beiden unabhängigen Variablen (siehe Tab. 10) zeigt, dass die Skala „MusikAllgemein“ bei weitem den größten Einfluss auf die abhängige Variable hat (Beta). Dies ist insofern erklärbar, als dass die Aufgaben in denselben Aufgabenblöcken, also zum Teil in denselben Sätzen vorkommen.

Tab. 10:
Koeffizienten der unabhängigen Variablen

	B	Standard- fehler	Beta	t	p	Δ
MusikAllgemein	.826	.045	.753	18.550	< .001	.776
Deutsch	.180	.043	.171	4.211	< .001	.013

Anmerkung: Δ = Erklärte Varianz pro Stufe

Mögliche Prädiktoren

Wie weiter oben erwähnt, wurden auch mögliche Prädiktoren getestet. Hierzu gehörten die letzte erinnerte Benotung im Fach Musik, die Einschätzung der Test-Schwierigkeiten durch die Fachlehrer der getesteten Klassen sowie die Berechnung des eher technischen Lesbarkeitsindex (= *LIX*, vgl. Lenhard, 2013) für alle Teilskalen. Leider erwiesen diese sich derzeit noch als wenig zuverlässige Werte: Die Validität eines retrospektiven Wertes wie der erinnerten Schulnote ist leider stark zu bezweifeln. Zusätzlich arbeiten einige Schulformen in der fünften und sechsten Klasse (noch) nicht mit dem Einsatz der Notenskala von 1 bis 6. Die Bewertung der Teil-Tests durch Fachlehrerinnen und Fachlehrer fiel weiterhin im Vergleich zu den realen Testergebnissen überwiegend zu schwierig und die Schätzungen des *LIX* meistens zu niedrig aus (vgl. Ahlers & Seifert, in Druck). Auch zur Bearbeitung dieses Punktes werden größere Stichproben benötigt, wenngleich hierin aber auch jetzt bereits prinzipiell Potenziale gesehen werden. Interessant wäre hier der Zusammenhang zwischen diesen Prädiktoren und den Gesamtfähigkeiten der Probanden.

4 Diskussion

Das beschriebene Instrument konnte sein Potenzial grundlegend innerhalb dieser Pilotierungsstudie beweisen. Dennoch ist es mit den vorhandenen Stichprobengrößen noch nicht möglich, valide Aus- oder Vorhersagen zu ermöglichen. Hierzu wäre eine größere, wünschenswerterweise dann bundesweite Stichprobenziehung notwendig. In deren Folge wären auch weitere mehrdimensionale Skalierungen (in denen jeder Test à 25 Aufgaben eine Dimension darstellt) möglich und sinnvoll. Dennoch sind die vorliegenden Ergebnisse und neuen Erkenntnisse vielversprechend und motivierend:

- a) Das Instrument basiert als erstes auf authentischen Texten und ist curricular sowie inhaltlich valide konstruiert. Sein Einsatz ist prinzipiell in allen Schulformen möglich, wenngleich in diesen Stichproben Hauptschüler bisweilen große Probleme in der Bearbeitung hatten.
- b) Die Reliabilitäten der Subskalen sind durchgängig gut bis sehr gut.

- c) Das dreidimensionale Modell bildet die spezifischen Teilkompetenzen im allgemeinen sprachlichen und (musik-)fachsprachlichen Kompetenzbereich sehr gut ab.
- d) Eine Bewertung der tatsächlichen sprachlichen und fachsprachlichen Kenntnisse und Kompetenzen ist für die gewählte Altersgruppe über den Einsatz des Instruments auf ökonomische Weise möglich.

Die Sonderstellung der Skalen zu musikalischen Formen – nämlich als sehr schwierig – ist nach Aussage der befragten Fachlehrer dadurch zu erklären, dass dieser inhaltliche Bereich bisweilen auch erst in der siebten Klasse unterrichtet wird, obwohl er bereits in den Bildungsplänen der fünften und sechsten Klasse zu finden ist. Dies gilt es, in Folgestudien zu berücksichtigen.

Künftig soll der spezifische Kompetenzbereich musikalischer Fachsprache und musikalischen Wissens näher fokussiert werden. Denkbar wäre der Bezug zu vorliegenden Teilkonstrukten musikalischer Kompetenz und den jeweiligen Instrumenten. Als langfristiges Ziel soll der Test auch handhabbar für den Einsatz in der pädagogischen Praxis gemacht werden. So könnte jede Lehrerin und jeder Lehrer in der siebten Klasse empirische Evidenz über die musiksprachlichen Kompetenzen und Wissensbestände der Schüler ihrer Klassen erhalten und ggf. entsprechende, individualisierte Interventionen planen, durchführen und evaluieren. Die „Fachsprache Musik“, ihr Umfang, ihre Struktur und Nutzung sowie die dafür benötigten Sprach-/Lesefähigkeiten und kognitiven Prozesse bieten als Forschungsgegenstände noch zahlreiche Desiderata, deren Bearbeitung zu verbesserten didaktischen Materialien und Testinstrumenten führen kann. Hiermit könnte die Forschung darüber hinaus zur Verbesserung im Umgang mit den vorhandenen Heterogenitäten an Schulen und über Schulen hinweg beitragen.

Literatur

- Ahlers, M. & Seifert, A. (in Druck). Sprachliche Heterogenität im Musikunterricht. In A. Niessen & J. Knigge, (Hrsg.), *Theoretische Rahmung und Theoriebildung in der musikpädagogischen Forschung* (Musikpädagogische Forschung, Bd. 36). Münster: Waxmann.
- Baur, R. S., Grotjahn, R. & Spettmann, M. (2006). Der C-Test als Instrument der Sprachstandserhebung und Sprachförderung im Bereich Deutsch als Zweitsprache. In J.-P. Timm (Hrsg.), *Fremdsprachenlernen und Fremdsprachenforschung* (S. 389–406). Tübingen: Narr.
- Becker-Mrotzek, M., Schramm, K., Thürmann, E. & Vollmer, H.-J. (Hrsg.). (2013). *Sprache im Fach. Sprachlichkeit und fachliches Lernen. Fachdidaktische Forschungen* (3. Aufl.). Münster: Waxmann.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Berlin: Springer. <http://doi.org/10.1007/978-3-540-33306-7>
- Deutsches PISA-Konsortium (Hrsg.). (2000). *PISA 20. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Gogoll, A. (2014). Das Modell der sport- und bewegungskulturellen Kompetenz und seine Implikationen für die Aufgabenkultur im Sportunterricht. In M. Pfizner (Hrsg.),

- Aufgaben im Sportunterricht – Konzepte und Forschungsbefunde zum kompetenzorientierten Sportunterricht.* Wiesbaden: VS-Verlag. Manuskript eingereicht zur Publikation.
- Grotjahn, R. (Hrsg.). (1994). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen.* Bochum: AKS.
- Grotjahn, R. (Hrsg.). (2002). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen.* Bochum: AKS.
- Harsch, C. & Hartig, J. (2010). Empirische und inhaltliche Analyse lokaler Abhängigkeiten im C-Test. In R. Grotjahn (Hrsg.), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-Test: Contributions from Current Research* (S. 205–232). Frankfurt/Main: Lang.
- Hasselhorn, J. & Lehmann, A. C. (2014). Entwicklung eines empirisch überprüfbaren Modells musikpraktischer Kompetenz (KOPRA-M). In B. Clausen (Hrsg.), *Teilhabe und Gerechtigkeit* (Musikpädagogische Forschung, Bd. 35, S. 77–93). Münster: Waxmann.
- Jank, W. (Hrsg.). (2005). *Musik-Didaktik. Praxishandbuch für die Sekundarstufe I und II.* Berlin: Cornelsen Scriptor.
- Jordan, A.-K. (2014). *Empirische Validierung eines Kompetenzmodells für das Fach Musik.* Münster: Waxmann.
- Jordan, A.-K., Knigge, J., Lehmann, A. C., Niessen, A. & Lehmann-Wermser, A. (2012). Entwicklung und Validierung eines Kompetenzmodells im Fach Musik – Wahrnehmen und Kontextualisieren von Musik. *Zeitschrift für Pädagogik*, 58(4), 500–521.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, P., Prenzel, M. et al. (2007). *Zur Entwicklung nationaler Bildungsstandards.* Bonn: Bundesministerium für Bildung und Forschung (BMBF). Zugriff am 16.04.2015. Verfügbar unter: http://www.bmbf.de/pub/zur_entwicklung_nationaler_bildungsstandards.pdf
- Knigge, J. (2011). *Modellbasierte Entwicklung und Analyse von Testaufgaben zur Erfassung der Kompetenz „Musik wahrnehmen und kontextualisieren“.* Berlin: LIT.
- Knigge, J. (2014). Der Kompetenzbegriff in der Musikpädagogik: Verwendung, Kritik, Perspektiven. In J. Vogt, M. Brenk & F. Heß (Hrsg.), *(Grund-)Begriffe musikpädagogischen Nachdenkens – Entstehung, Bedeutung, Gebrauch.* Münster: LIT. Manuskript eingereicht zur Publikation.
- Kopiez, R. & Wolf, A. (2014, Oktober). *Zur Entwicklung eines kompetenzbasierten Gehörbildungstests auf Hochschulniveau.* Präsentation auf der Jahrestagung des Arbeitskreis für Musikpädagogische Forschung am, Neuwied Engers.
- Lenhard, W. (2013). *Leseverständnis und Lesekompetenz: Grundlagen – Diagnostik – Förderung.* Stuttgart: Kohlhammer.
- Mashkovskaya, A. (2013). *Der C-Test als Lesetest bei Muttersprachlern.* Dissertation, Universität Duisburg-Essen. Zugriff am 16.04.2015. Verfügbar unter <http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=32859>
- Niessen, A. & Knigge, J. (Hrsg.). (in Druck). *Theoretische Rahmung und Theoriebildung in der musikpädagogischen Forschung.* (Musikpädagogische Forschung, Bd. 36). Münster: Waxmann.
- Niessen, A., Lehmann-Wermser, A., Knigge, J. & Lehmann, A. C. (2008). Entwurf eines Kompetenzmodells „Musik wahrnehmen und kontextualisieren“. *Zeitschrift für Kritische Musikpädagogik [Sonderedition: Bildungsstandards und Kompetenzmodelle für das Fach Musik?]*, S. 3–33. Zugriff am 16.04.2015. Verfügbar unter <http://www.zfkm.org/sonder08-niessenetal.pdf>
- Nold, G. & Willenberg, H. (2007). Lesefähigkeit [Englisch und Deutsch]. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messungen. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 23–41). Weinheim: Beltz.

- Rautenberg, I. (2012). *Musik und Sprache: eine Längsschnittstudie zu Effekten musikalischer Förderung auf die schriftsprachlichen Leistungen von GrundschülerInnen*. Baltmannsweiler: Schneider-Verlag Hohengehren.
- Rost, J. (1996). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- von Davier, A. A., Carstensen, C. H. & von Davier, M. (2006). *Linking competencies in educational settings and measuring growth (No. RR-06–12)*. Princeton, NJ: Educational Testing Service.
- von Georgi, R. (2010). Test. In H. de la Motte- Haber, H. von Loetsch, G. Rötter & C. Utz (Hrsg.), *Lexikon der Systematischen Musikwissenschaft* (S. 480–483). Laaber: Laaber.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim: Beltz.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1997). *ConQuest: Multi-Aspect Test Software* [Computer program]. Camberwell, AU: Australian Council for Educational Research.