

Vol. 4, No. 2  
Oktober 2013

ISSN: 2190-3174

Andreas Lehmann-Wermser (Hrsg./ed.)

Elektronischer Artikel:

**Sonja Nonte**

*Georg-August-Universität Göttingen*

**Herausforderungen und Probleme bei der Entwicklung  
eines Instruments zur Selbsteinschätzung  
musikalischer Fähigkeiten im Grundschulalter**

**Challenges and problems for the development of an instrument  
for the self-evaluation of musical abilities in elementary school age**

Elektronische Version:

<http://www.b-em.info/index.php?journal=ojs&page=article&op=view&path%5B%5D=94&path%5B%5D=248>

[urn:nbn:de:101:1-201312191344](http://nbn-resolving.org/urn:nbn:de:101:1-201312191344)

© Sonja Nonte 2013 All rights reserved

## **Herausforderungen und Probleme bei der Entwicklung eines Instruments zur Selbsteinschätzung musikalischer Fähigkeiten im Grundschulalter**

*Sonja Nonte*

### **Abstract**

Auf der Grundlage von theoretischen und empirischen Vorarbeiten u. a. von Marsh (1992), Vispoel (1993) und Kröner, Schwanzer und Dickhäuser (2009) wurde ein Instrument zur Erfassung des musikalischen Selbstkonzepts von Grundschulkindern mit den zwei Facetten ‚Singen‘ und ‚Musizieren‘ entwickelt. In einer Pilotstudie ( $n = 88$ ) sowie in einer längsschnittlich angelegten Evaluationsstudie des Programms „Jedem Kind ein Instrument“ (JeKi) mit drei Messzeitpunkten ( $n = 826$ ) wurde das Instrument auf die gängigen Gütekriterien hin überprüft. Die theoretisch angenommene Struktur des Instruments wurde anhand von Faktorenanalysen überprüft und bestätigt. Korrelationen zu Außenkriterien bekräftigen zudem die Validität des Instruments. Retest-Reliabilitäten nehmen über die Zeit zu und betragen über den Zeitraum von etwa einem Jahr (vom zweiten zum dritten Schuljahr)  $r = .61$  für die Facette ‚Singen‘. Des Weiteren wurden Überprüfungen auf Messäquivalenz im Hinblick auf die Zeit, das Geschlecht und für den Gruppierungsfaktor – Kinder mit Erfahrung im Instrumentalunterricht gegenüber Kindern ohne – vorgenommen. Für das vorliegende Instrument kann metrische statistische Invarianz über die Zeit sowie für den Gruppierungsfaktor Instrumentalunterricht und mindestens strenge statistische Invarianz für das Geschlecht beobachtet werden. Geschlechts- und Gruppenunterschiede deuten auf reziproke Effekte sozialer und dimensionaler Vergleichsprozesse hin, können jedoch anhand der vorliegenden Studie nicht weiter betrachtet werden.

**Schlagwörter:** Geschlechtsrollenstereotype, Instrumentalunterricht, Längsschnittstudie, Selbstkonzept

## Summary

On the basis of previous theoretical and empirical works by i. a. Marsh (1992), Vispoel (1993) and Kröner, Schwanzer and Dickhäuser (2009) an instrument for the assessment of the musical self-concept with the two subscales 'singing' and 'making music' of primary school pupils was developed. The instrument was tested for validity during a pilot study ( $n = 88$ ) and a longitudinal study, which is part of the JeKi-program ("For each child an instrument" [Jedem Kind ein Instrument]), whose three measurements have been evaluated ( $n = 826$ ). Factor analysis provided support for the theoretical assumptions of the structure of the scales. Retest-reliabilities increased during a period of three years. For the subscale singing a retest-reliability of  $r = .61$  for the self-concept in the second and third class could be reported. Correlations between the scales and the external criteria support the validity of the instrument. Furthermore measurement invariance was checked for the three underlying measurement points as well as for gender differences and the grouping factor with children who are experienced in instrumental lessons and children who are not. There is a minimum of metric statistical invariance for the different time points and the grouping factor instrumental lessons and a minimum of strong statistical invariance for the grouping factor gender. Mean-differences between these groups indicate reciprocal effects of social and dimensional comparisons, but couldn't be explored on the basis of this study.

**Keywords:** gender, instrumental instruction, longitudinal study, self-concept

## 1. Einleitung

Obgleich die standardisierte Befragung von Kindern als Expertinnen und Experten ihrer Umwelt in vielen Studiendesigns vorgesehen ist (Borgers, de Leeuw & Hox, 2000), so fehlt es doch häufig an elaborierten und ausreichend erprobten Verfahren, die zudem noch in längsschnittlichen Studien eingesetzt werden können (Rauer & Schuck, 2004). Im vorliegenden Fall wird die Entwicklung und Erprobung eines standardisierten Erhebungsinstruments vorgestellt. Das Ziel der Arbeit besteht darin, die Entwicklung eines Instruments zur Erfassung des musikalischen Fähigkeitsselbstkonzepts im Grundschulalter vorzustellen und die Güte des Instruments anhand einer Stichprobe aus der längsschnittlich angelegten „Studie zum Instrumentalunterricht in Grundschulen“ (SIGrun) zu überprüfen.

## 2. Theoretischer Hintergrund

Das Konstrukt des musikalischen Fähigkeitsselbstkonzepts ist Bestandteil eines hierarchisch strukturierten multidimensionalen Mehrfacetten-Modells, bei dem das generelle Selbstkonzept als globales latentes Konstrukt die Wahrnehmung einer Person über sich selbst beinhaltet (Shavelson, Hubner & Stanton, 1976). Diese Wahrnehmung wird durch die Interaktion mit der Umwelt und besonders durch die Interaktion mit „significant others“ wie z. B. Eltern oder Peers beeinflusst (Shavelson et al., 1976, S. 411). Das generelle Selbstkonzept gliedert sich in die Bereiche des akademischen und nicht-akademischen Selbstkonzepts. Marsh, Byrne und Shavelson (1988) konnten weiterhin zeigen, dass sich die Selbstkonzeptfacetten hinsichtlich eines mathematischen und eines verbalen Selbstkonzepts unterscheiden lassen. Somit differenziert sich das globale Selbstkonzept immer weiter in spezifische Unterfacetten des Selbstkonzepts aus. Bei jungen Kindern können die Facetten in der Regel weniger differenziert erfasst werden als bei älteren Kindern (Harter & Pike, 1984).

Vispoel (1995) überprüfte das Vorliegen eines eigenständigen Faktors des musikalischen Selbstkonzepts anhand einer Stichprobe von  $n = 831$  Studierenden mit Hilfe von konfirmatorischen Faktorenanalysen. Zur Erfassung der unterschiedlichen Fähigkeitsselbstkonzepte verwendete er die Items des „Self Description Questionnaire“ (SDQ-III; Marsh, 1987) und des „Arts Self-Perception Inventory“ (ASPI; Vispoel, 1993). Die Befunde sprechen dabei für die Annahme des Modells von Marsh (1990), wonach neben dem akademisch/verbalen, akademisch/nonverbalen und nicht-akademischen Selbstkonzept ein weiterer Faktor, das musische Fähigkeitsselbstkonzept (mit den Facetten Theater, Tanz, Kunst, Musik) beobachtet werden kann.

Im internationalen Bereich formulierte Marsh (1987) im Kontext des SDQ-III einzelne Items zur Erfassung der Wahrnehmung eigener künstlerischer Fähigkeiten. Vispoel (1993) entwickelte auf der Grundlage von Vorarbeiten von Harter (1988), Marsh (1990) und anderen ein Instrument zur Erfassung des Selbstkonzepts von Jugendlichen sowie eine Form für Erwachsene in den vier künstlerischen Bereichen Musik, Kunst, Tanz und Theater. Die zwei Instrumente umfassen je Subskala 10 Items auf einer vier- bzw. achtstufigen Skala. Die Ska-

len weisen hohe interne Konsistenzen und geringe Überschneidungen zwischen den einzelnen Facetten auf.

Bisher liegen im deutschsprachigen Raum nur wenige Arbeiten zur Erfassung des Fähigkeitsselbstkonzepts im musikalischen und künstlerischen Bereich vor (Schwanzer, 2002; Pfeiffer, 2007; Spychiger, 2007). Eine der ersten publizierten empirischen Forschungsarbeiten in Deutschland zum Thema Fähigkeitsselbstkonzept im musisch-künstlerischen Bereich stellt die Arbeit von Schwanzer (2002) dar. Schwanzer entwickelte ein Instrument zur Erfassung des musikalisch-künstlerischen Selbstkonzepts junger Erwachsener auf Grundlage der Arbeiten von Vispoel (1993, 1995, 1996) und Marsh (1990). Insgesamt umfasst die Skala des Selbstkonzepts musikalisch-künstlerischer Fähigkeiten 17 Items, die affektive („Ich mag es, mich durch künstlerische oder musikalische Aktivitäten zu entfalten“) und kognitive („Mit Kunst oder Musik kenne ich mich gut aus“) Komponenten beinhaltet. Damit erfasst das Instrument die Selbstwahrnehmung musikalisch-künstlerischer Fähigkeiten im Allgemeinen, unterscheidet aber nicht zwischen Schauspiel, Malerei, Singen und anderen Facetten. Obschon das Instrument eine hohe interne Konsistenz (Cronbachs Alpha = .93) aufweist, muss die inhaltliche Validität der Items aufgrund fehlender inhaltlicher Abgrenzung angezweifelt werden. So scheint es möglich, dass eine Aussage für den künstlerischen Bereich zutreffend ist, aber für den musikalischen Bereich keine Gültigkeit besitzt. Eine Trennung musikalischer Fähigkeitseinschätzungen und anderer künstlerischer Fähigkeiten scheint dementsprechend notwendig zu sein. In der Musikpädagogik stark rezipierte Arbeiten zur Erfassung des musikalischen Selbstkonzepts stellen die Arbeiten von Spychiger (2007) sowie Spychiger, Grube und Olbertz (o. J.) dar. Der bewusst gewählte explorativ-qualitativ Zugang zum Forschungsfeld grenzt sich methodisch und inhaltlich von den zuvor genannten Arbeiten ab. Im Kontext dieser Arbeiten wird das musikalische Selbstkonzept als „Korrelat akkumulierter musikalischer Erfahrung, anhand derer eine Person sich in diesem Lebensbereich selbst wahrnimmt und einschätzt und worauf sie rekurriert, wenn sie sich diesbezüglich beschreibt“ (Spychiger et al., o. J., S. 14) betrachtet. Eine reliable und valide Einschätzung eigener Fähigkeiten kann, insbesondere bei Kindern ohne eine vertiefende musikalische Vorprägung, vermutlich nicht vorgenommen werden. Hier kann ein quantitatives Verfahren genutzt werden, um diese Limitationen zu klären und zu beheben. Pfeiffer (2007) überprüfte anhand einer großangelegten Studie mit  $n = 2400$  Schülerinnen und Schülern der fünften bis dreizehnten Jahrgangsstufen an Gymnasien ein selbstentwickeltes Instrument zur Erfassung des Selbstkonzepts musikalischer Fähigkeiten. Die Items wurden in Anlehnung an die Arbeiten Vispoels (1993, 1995, 1996) formuliert sowie um die Facetten der Freude beim Umgang mit Musik, Bedeutsamkeit von Musik, Kommunikation durch Musik, Sich-Sonnen in einem Chor bzw. Ensemble und das Persönlichkeitsmerkmal Extraversion erweitert und ergänzt. Angaben zur Skalengüte wurden bisher noch nicht publiziert. Einschränkend muss gesagt werden, dass die Stichprobe auf das Alter bezogen stark heterogen ist. Eine Unterscheidung zwischen den Altersgruppen oder hinsichtlich der Geschlechtszugehörigkeit wird nicht vorgenommen. Es scheint jedoch vorstellbar, dass sich die selbstbezogenen Fähigkeitseinschät-

zungen diesbezüglich unterscheiden (Schwanzer, Trautwein, Lüdtke & Sydow, 2005; Eccles, Wigfield, Harold & Blumenfeld, 1993).

Auswirkungen des musikalischen Selbstkonzepts auf das Ausmaß aktiver musikalischer Partizipation sowie Auswirkungen des Schulprofils auf das musikalische Selbstkonzept wurden in einer Explorationsstudie von Fritzsche, Kröner und Pfeiffer (2011) in den Blick genommen. Das musikalische Selbstkonzept wurde anhand einer fünf Item starken Skala in Anlehnung an Schwanzer et al. (2005) erfasst. Die Gesamtstichprobe wurde in drei Teilstichproben aufgeteilt, die ausgeprägte musische Gymnasien besuchten sowie andere Gymnasien ohne musikalisches Profil. Die Kinder waren durchschnittlich  $M = 14.20$  Jahre ( $SD = 2.32$ ) alt. Fritzsche et al. (2011) fanden keinen Unterschied im musikalischen Selbstkonzept zwischen Schülerinnen und Schülern die ein Gymnasium mit einem ausgeprägten musikalischen Profil besuchten im Vergleich zu Schülerinnen und Schülern von anderen Schulen. Es zeigten sich keine substantiellen Effekte des musikalischen Selbstkonzepts auf musikalische Aktivitäten. Einschränkungen der Befunde müssen nach eigenen Angaben aufgrund der hohen Eingangsselektivität der Stichprobe und des querschnittlichen Designs hingenommen werden. Möglicherweise gehen die geringen oder nicht beobachtbaren Effekte zu Lasten des verwendeten Instruments in Anlehnung an Schwanzer et al. (2005). Arbeiten zum musikalischen Selbstkonzept konzentrierten sich demnach bisher auf die Entwicklung von Instrumenten zur Erfassung des musikalischen Selbstkonzepts von Heranwachsenden im Sekundarbereich (Pfeiffer, 2007; Bernecker, Haag & Pfeiffer, 2006; Schwanzer et al., 2005; Vispoel, 1993) sowie im Erwachsenenalter (Schwanzer, 2002; Vispoel, 1995, 1996). Die Erfassung des musikalischen Selbstkonzepts bei Kindern im Grundschulalter stand bisher kaum im Zentrum von empirischen Forschungsarbeiten (Kröner, Schwanzer & Dickhäuser, 2009). Kröner et al. (2009) legten eine deutsche Fassung zur Ergänzung des SDQ-III von Marsh (1992) vor, die eine vier Item starke Skala zur Erfassung des multidimensionalen Selbstkonzepts in Hinblick auf musikalische Fähigkeiten umfasst. Insgesamt konzentrierten sich ihre Untersuchungen auf eine Stichprobe von Grundschulkindern ( $n = 166$ ) in einer altersmäßig stark heterogenen Gruppe (zweite, dritte und vierte Jahrgangsstufe). Die interne Konsistenz der verwendeten Skala „musikalisches Selbstkonzept“ fällt mit Cronbachs Alpha = .61 vergleichsweise niedrig aus. Hier wäre eine Verlängerung des Instruments um eine größere Anzahl an Items sowie eine altersmäßig homogenere Stichprobenzusammensetzung vermutlich hilfreich.

Die aufgeführten Arbeiten zur Entwicklung eines Instruments zur Erfassung des musikalischen Selbstkonzepts im Kindesalter verdeutlichen das Forschungsdesiderat in diesem Bereich. Es fehlt ein valides und reliables Instrument zur Erfassung des musikalischen Selbstkonzepts im Grundschulalter. Die Grundlage bildet dabei die Annahme, dass bereits Kinder im Grundschulalter über ein hierarchisches, mehrdimensionales Fähigkeitsselbstkonzept verfügen, bei dem sich einzelne Facetten eindeutig voneinander trennen lassen (Marsh, Barnes, Cairns & Tidman, 1984). Das Ziel dieser Arbeit besteht darin, die Entwicklung eines solchen Instruments vorzustellen. Dabei sollte das Instrument möglichst ökonomisch im

Gruppenkontext einsetzbar und über die Laufzeit der Studie von insgesamt vier Jahren statistisch invariant sein. So kann angenommen werden, dass die noch fehlende Stabilität des Selbstkonzepts im Kindesalter für Item- und Parametervariationen über die Zeit verantwortlich sein kann (Harter, 1982, Harter & Pike, 1984). Im Fokus der Instrumentenentwicklung stehen insbesondere die beiden Bereiche Singen und Musizieren (ein Instrument spielen). Hinweise auf die Validität liefern Hintergrundvariablen zur Freude am Musizieren und Singen sowie Einschätzungen der musikalischen Fähigkeiten durch relevante Bezugspersonen, die Eltern und Lehrkräften. Des Weiteren sollen erste Mittelwertvergleiche Aufschluss über mögliche Geschlechtsunterschiede bei der Einschätzung musikalischer Fähigkeiten geben. U. a. Eccles et al. (1993) sowie Busch & Kranefeld (2012) konnten beobachten, dass Jungen sich hinsichtlich ihrer musikalischen Fähigkeiten geringer einschätzen als Mädchen. Eccles et al. (1993) sowie Fritzsche et al. (2012) fanden Unterschiede in der Ausprägung des musikalischen Selbstkonzepts in Abhängigkeit eigener Erfahrungen. Anhand von ersten Mittelwertvergleichen sollen zudem Unterschiede zwischen Kindern, die Instrumentalunterricht erhalten und Kindern, die keinen Instrumentalunterricht erhalten, hinsichtlich ihres musikalischen Selbstkonzepts berichtet werden.

### 3. Methode

#### 3.1 Stichprobe

Die im Kontext dieser Arbeit genutzten Daten stammen aus einem Projekt eines bundesweit angelegten Forschungsverbundes. Es handelt sich dabei um die Studie zur Evaluation des in Grundschulen im Ruhrgebiet und in Hamburg angesiedelten Programms „Jedem Kind ein Instrument“ (JeKi).<sup>1</sup> Die finanziellen Mittel für die Begleitevaluation zu diesem Programm werden vom BMBF im Rahmen der Förderrichtlinie „Empirische Bildungsforschung“ bereitgestellt.

Das Teilprojekt „Transfer“ aus der „Studie zum Instrumentalunterricht in Grundschulen“ (SIGrun) nimmt den Einfluss von aktivem Musizieren auf die Entwicklung von Einstellungen, Interessen und auf das schulische Lernverhalten der Schülerinnen und Schüler in den Blick. Die Studie ist längsschnittlich angelegt, wobei jeweils zum Ende des Schuljahres Schülerinnen und Schüler, Eltern sowie die schulischen Akteure mittels standardisierter Verfahren im Klassenkontext befragt werden. Die Stichprobenziehung erfolgte stratifiziert und anhand eines systematischen Samplings (Listenziehung). Sie umfasst Kinder an Schulen, die an dem JeKi Programm in Nordrhein-Westfalen teilnehmen ( $n = 255$ ), Kinder an Schulen, die an dem JeKi Programm in Hamburg ( $n = 281$ ) teilnehmen, Kinder an Schulen, die in Hamburg auf bereits langjährige Erfahrungen im Instrumentalunterricht zurückblicken können ( $n = 70$ ) und Kinder an Schulen, die an dem Programm „bewegte Schule“ in Nordrhein-

---

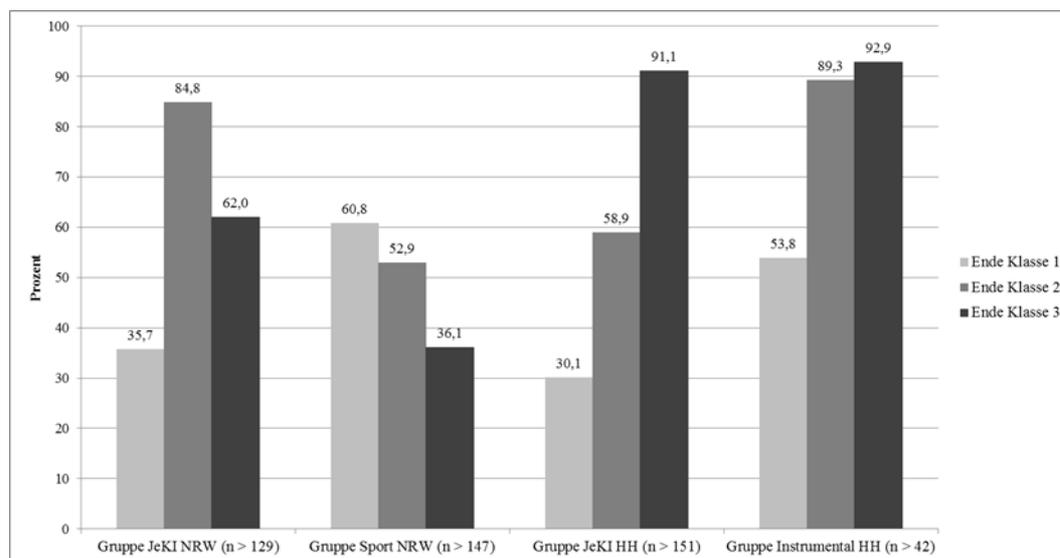
<sup>1</sup> Mehr Informationen zum Programm unter <http://www.jedemkind.de/> [12.11.2012] sowie unter <http://www.hamburg.de/jeki/> [12.11.2012].

Westfalen teilnehmen ( $n = 211$ ). Zum ersten Messzeitpunkt am Ende des ersten Schuljahres 2008/2009 umfasst die Stichprobe  $n_1 = 1112$  Schülerinnen und Schüler mit einem durchschnittlichen Alter von  $M = 7.44$  ( $SD = 0.48$ ). Es beteiligten sich  $n = 440$  Mädchen und  $n = 377$  Jungen an der Befragung. Zum zweiten Messzeitpunkt am Ende der zweiten Klasse im Schuljahr 2009/2010 kam es aufgrund von Ausfällen von einzelnen Schulen und einer daraus resultierenden Aufstockung zu einer Stichprobenverschiebung. Es wurden  $n_2 = 1243$  Schülerinnen und Schüler mit einem durchschnittlichen Alter von  $M = 8.33$  ( $SD = 0.43$ ) befragt. Im dritten Schuljahr (2010/2011) nahmen erneut 1144 Schülerinnen und Schüler mit einem durchschnittlichen Alter von  $M = 9.33$  ( $SD = 0.51$ ) an der Studie teil. Insgesamt beteiligten sich 72 Klassen in 28 Grundschulen in Hamburg und Nordrhein-Westfalen an den Befragungen. Eine Schule mit drei Klassen entschied sich aufgrund des aus ihrer Sicht hohen Aufwandes nach dem ersten Messzeitpunkt gegen eine erneute Befragung.<sup>2</sup> Der Anteil an Kindern die zu mindestens einem Messzeitpunkt Instrumentalunterricht erhalten haben, beträgt 75,1 Prozent ( $n = 550$ ). Dementsprechend erhielten 24,9 Prozent ( $n = 182$ ) der Kinder zu den befragten Zeitpunkten keinen Instrumentalunterricht. Für 94 Kinder liegen diesbezüglich keine Angaben vor. Der Instrumentalunterricht erfolgt für fast die Hälfte der Kinder ( $n = 239$ ) über das Programm JeKi. Die Kinder erhielten verbindlich im Klassenverbund (in Hamburg) bzw. freiwillig im Anschluss an den regulären Unterricht (Nordrhein-Westfalen) Instrumentalunterricht auf einem selbstgewählten Instrument. Für alle anderen Fälle kann nicht trennscharf ermittelt werden, ob der Instrumentalunterricht im schulischen Setting oder privat erteilt wurde. Angaben zum Instrumentalunterricht nach Gruppen können Abbildung 1 entnommen werden.

---

<sup>2</sup> Angaben zu Dropout-Quoten können der Publikation von Nonte & Schwippert (2012) entnommen werden.

Abb. 1: Prozentuale Häufigkeiten des Merkmals „Kind erhält derzeit Instrumentalunterricht“ nach Angaben der Eltern von der ersten bis zur vierten Klasse über die Gruppen.



In die folgenden Analysen gehen Fälle ein, die zu allen drei Messzeitpunkten gültige Daten aufweisen. Dieser Datensatz umfasst Angaben von  $n_G = 826$  Schülerinnen und Schülern, die an allen drei Befragungszeiträumen teilgenommen haben.

### 3.2 Verfahren

*Musikalisches Selbstkonzept.* In einem ersten Schritt zur Konzeption eines Instruments zur Erfassung des musikalischen Selbstkonzepts wurden in Anlehnung an Vorarbeiten von Marsh (1992), Vispoel (1993) und Kröner et al. (2009) 21 Items formuliert, die auf die Einschätzung musikalischer Fähigkeiten aus der Perspektive der Kinder abzielen. Inhaltlich sollten die Items die Bereiche des Singens, Musizierens und Musikhörens sowie allgemeine Aspekte des musikalischen Selbstkonzepts abdecken. Dabei wurde bei der Formulierung der Items auf eine kindgerechte Sprache geachtet sowie auf die negative Formulierung von Items verzichtet. Marsh (1986) und Benson & Hocevar (1985) konnten anhand von Studien zeigen, dass die Formulierung von negativen Items häufig zu einer Ausbildung eines inhaltlich nicht interpretierbaren methodischen Artefakts führt. Anhand einer Pilotstudie mit  $n = 88$  Schülerinnen und Schülern aus vier ersten Klassen, die unabhängig von der Hauptstudie ermittelt werden konnten, wurde zunächst die Verständlichkeit der Items überprüft. Den Kindern wurden die einzelnen Items im Klassenkontext laut vorgelesen. Daraufhin wurden sie instruiert, jede ihrer Antworten auf einer vierstufigen Smileyskala einzuordnen (1 = ja, ganz genau bis 4 = nein, überhaupt nicht). Die eindeutige Zuordnung der Items zu den Smileyskalen wurde über Zuordnungssymbole gewährleistet. Mit Hilfe der Daten aus der Pilotstudie konnten anhand einer Hauptkomponentenanalyse mit obliquem Rotationsverfahren (Promax) erste Items eliminiert werden, die zu geringe Faktorladungen ( $<.30$ ) (Kline, 1997) aufwiesen oder zu Verständnisproblemen führten (vgl. Anhang 1 & 2). Dies betraf vor allem allgemein formulierte Items wie „Neue Lieder lerne ich schnell“ oder „Ich bin musika-

lich“ und zum anderen Items zum Musikhören. Am Ende verblieben 10 Items im Itempool. Die Auswahl erfolgte zunächst aufgrund statistischer Kennwerte. Die endgültig beibehaltenen Items beinhalten domänenspezifische Einschätzungen in den Bereichen Singen (Beispielitem: „Ich bin gut im Singen“) und Musizieren (Beispielitem: „Musikmachen fällt mir leicht“). Der Itempool umfasst zum einen Items, die die wahrgenommene Freude in Bezug auf die jeweilige musikalische Tätigkeit abbilden, zum anderen Items, die soziale Vergleiche im Klassenkontext thematisieren. Insgesamt wurden drei Items zur Erfassung von affektiven Komponenten und drei Items zur Erfassung von kognitiven Komponenten des Fähigkeits-selbstkonzepts Singen beibehalten. Die vier Items zur Fähigkeitsselbsteinschätzung im Musizieren beinhalten lediglich kognitive Komponenten (vgl. Abb. 2). Eine Aufstockung des Itempools mit neuen Items konnte aus zeitlichen und forschungspragmatischen Gründen nicht erfolgen. Die Befragung der Hauptstudie fand jeweils zum Ende eines Schuljahres während der regulären Unterrichtszeit in der Klasse statt. Um ein standardisiertes Vorgehen zu gewährleisten, wurden für die Erhebungen Testleiterinnen und Testleiter eingesetzt, die zuvor an einer Schulung teilgenommen hatten. Sie erhielten die Instruktion, sich exakt an das Testleitermanual zu halten und keine Verständnisfragen zu beantworten. Der Fragebogen wurde in einem kindlich ansprechenden Design vorgelegt und die einzelnen Fragen und Items laut vorgelesen. Die Zuordnung zum jeweiligen Item erfolgte auch hier durch Bilder.

Neben der Befragung von Schülerinnen und Schülern wurden zusätzlich zu jedem Messzeitpunkt auch die Eltern mit einem standardisierten Fragebogen befragt. Neben soziodemographischen Hintergrundvariablen wurden musik- und sportspezifische Einstellungen sowie die kulturelle Praxis der Familie erfasst, beispielsweise die Freude des Kindes am Singen und die Musikalität des Kindes aus Sicht der Eltern. Zusätzlich zu den Angaben der Eltern wurden auch die Lehrkräfte um eine Einschätzung der musikalischen Fähigkeiten der Schülerinnen und Schüler gebeten.

*Freude am Singen.* Die Freude des Kindes am Singen wurde im Elternfragebogen zum ersten, zweiten und dritten Messzeitpunkt auf einer 5-stufigen Skala als Einzelitem erfasst (1 = gar nicht gerne, 2 = nicht gerne, 3 = weder gerne noch ungerne, 4 = gerne, 5 = sehr gerne). Die zugehörige Frage lautete „Wie gerne singt ihr Kind?“ (Busch & Wagschal, 2010). Es zeigen sich folgende Zusammenhänge zwischen den Messzeitpunkten: Für den ersten und zweiten Messzeitpunkt korrelieren die Einschätzungen der Eltern zu  $r = .60$  ( $p < .01$ ,  $n = 437$ ), für den zweiten und dritten Messzeitpunkt zu  $r = .65$  ( $p < .01$ ,  $n = 374$ ) und für den ersten und letzten Messzeitpunkt zu  $r = .52$  ( $p < .01$ ,  $n = 443$ ).

*Musikalische Fähigkeiten.* Die musikalischen Fähigkeiten wurden zu allen drei Messzeitpunkten über die Lehrkräfte an der Schule mit der Frage „Wie schätzen Sie die musikalischen Fähigkeiten des Schülers/der Schülerin ein?“ erfragt. Die Antworten konnten auf einer 5-stufigen Skala (1 = sehr gut, 2 = gut, 3 = mittel, 4 = schlecht, 5 = sehr schlecht) abgetragen werden. Die Stabilität (Korrelation nach Spearman) des Lehrkrafturteils beträgt  $r = .59$  ( $p < .01$ ,  $n = 357$ ) für den ersten und zweiten Messzeitpunkt,  $r = .64$  ( $p < .01$ ,  $n = 454$ ) für den zweiten und dritten Messzeitpunkt und  $r = .46$  ( $p < .01$ ,  $n = 420$ ) für den ersten und dritten

Messzeitpunkt. Weiterhin wurden die Eltern gefragt, für wie musikalisch sie ihr Kind halten. Die Antwortkategorien reichen von 1 bis 5 (1 = unmusikalisch, 2 = eher unmusikalisch, 3 = weder unmusikalisch noch musikalisch, 4 = eher musikalisch, 5 = sehr musikalisch). Hier betragen die Zusammenhänge (Korrelation nach Spearman) für den ersten und zweiten Messzeitpunkt  $r = .57$  ( $p < .01$ ,  $n = 373$ ), für den zweiten und dritten  $r = .62$  ( $p < .01$ ,  $n = 342$ ) und für den ersten und dritten Messzeitpunkt  $r = .49$  ( $p < .01$ ,  $n = 372$ ).

*Instrumentalunterricht.* Die Eltern wurden zu jedem Messzeitpunkt gefragt „Erhält Ihr Kind derzeit Unterricht auf einem Musikinstrument?“. Die Antwortmöglichkeiten waren „Ja, auf einem“, „Ja, auf mehreren“ und „Nein“. Für die Auswertungen wurde jeweils eine dichotome Variable erzeugt (0 = Nein, 1 = Ja). Zum ersten Messzeitpunkt liegen gültige Angaben von  $n_1 = 624$ , zum zweiten Messzeitpunkt von  $n_2 = 510$  und zum dritten Messzeitpunkt von  $n_3 = 500$  Eltern vor.

### 3.3 Statistisches Vorgehen

Um die hypothetisch postulierte Struktur des musikalischen Selbstkonzepts mit den Facetten Singen und Musizieren einer empirischen Überprüfung zu unterziehen, wurden konfirmatorische Faktorenanalysen durchgeführt. Hinweise auf die konvergente Validität des Instruments liefern Zusammenhangsanalysen der zwei Subfacetten mit den Einschätzungen zu den musikalischen Fähigkeiten durch die Grundschullehrkraft und die Eltern. Die Überprüfung auf Messinvarianz über die Zeit und über Gruppierungsfaktoren (Geschlecht und Erfahrung mit Instrumentalunterricht) erfolgt anhand von konfirmatorischen Faktorenanalysen. Die Berechnungen werden mit der Software *Mplus 6.1* (Muthén & Muthén, 2010) durchgeführt. Da es sich bei der vorliegenden Stichprobe um geschachtelte Daten handelt (Schülerinnen und Schüler in Klassen), wurde in *Mplus* eine Korrektur des Standardfehlers durch den Befehl *type=complex* vorgenommen. Die Klassenzugehörigkeit bildete als kleinste Aggregateneinheit die Clustervariable. Um zudem der ordinal/kategorialen Datenstruktur gerecht zu werden, wurde ein praktikabler und für ordinale Datenstrukturen entwickelter Schätzer (*WLSMV*) verwendet (Muthén, du Toit & Spisic, 1997). Fehlende Werte lagen kaum vor ( $< 5\%$ ) und wurden mit der Standardeinstellung von *Mplus* durch das Maximum Likelihoodverfahren (*FIML*) berücksichtigt (Lüdtke, Robitzsch, Trautwein & Köller, 2007).

## 4. Ergebnisse

### 4.1 Faktorielle Validität

Die Überprüfung der Konstruktvalidität der zwei Subdimensionen Musizieren und Singen erfolgt anhand von konfirmatorischen Faktorenanalysen. Das theoretisch postulierte Faktorenmodell mit den zwei Facetten Singen und Musizieren wurde empirisch gegen ein Generalfaktorenmodell und ein Faktorenmodell mit einem genesteten Faktor getestet (vgl. Tabelle 1). Das genestete Faktorenmodell zeichnet sich dadurch aus, dass zusätzliche Varianz der einzelnen Items neben den zwei Facetten durch einen dritten, globalen latenten Faktor erklärt wird. Zur Beurteilung der Modellgüte wurden der *Root Mean Square Error of Approximation*

on (RMSEA), der *Tucker-Lewis-Index* (TLI) und der *Comparative Fit Index* (CFI) herangezogen. Werte des *RMSEA* unter .06 sowie des TLI und des CFI über .90 weisen auf eine gute Modellanpassung hin (Hu & Bentler, 1999).

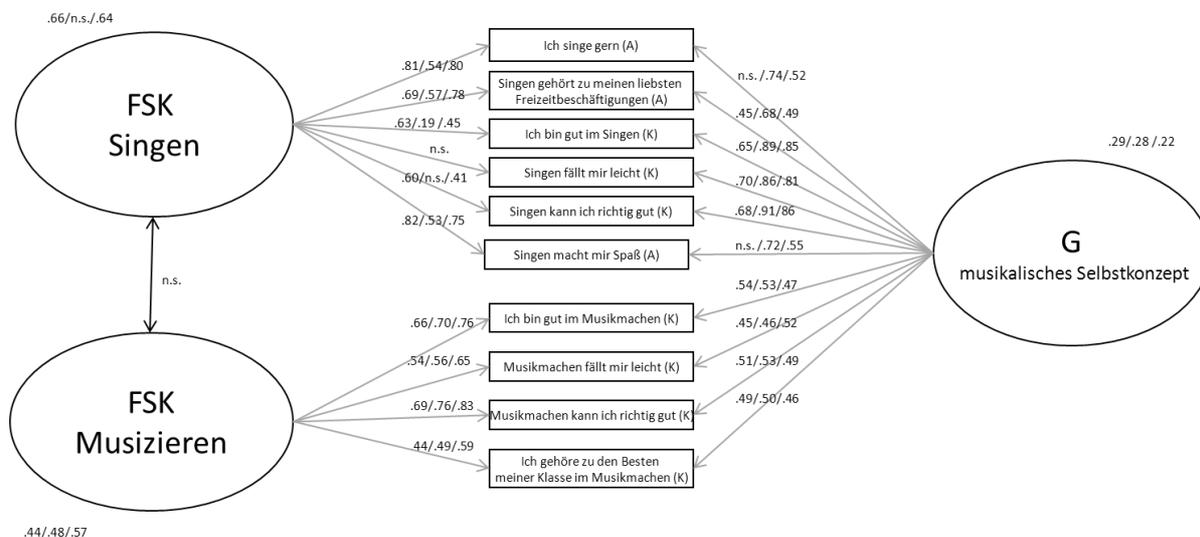
Tab. 1: Modelltestung anhand von konfirmatorischen Faktorenanalysen zu drei Messzeitpunkten.

		$\chi^2$	<i>df</i>	<i>p</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>
Generalfaktoren-Modell	MZP1	709.099	35	.000	.906	.880	.156
	MZP2	1186.704	35	.000	.908	.882	.200
	MZP3	1681.594	35	.000	.927	.906	.239
2-Facetten-Modell	MZP1	135.686	34	.000	.986	.981	.061
	MZP2	307.593	34	.000	.978	.971	.099
	MZP3	408.415	34	.000	.983	.978	.115
genestetes Faktorenmodell	MZP1	63.302	24	.000	.995	.990	.045
	MZP2	84.087	24	.000	.995	.991	.055
	MZP3	107.586	24	.000	.996	.993	.065

Anmerkungen: MZP1 = Messzeitpunkt 1 (erste Klasse; n = 794), MZP2 = Messzeitpunkt zwei (zweite Klasse; n = 826), MZP3 = Messzeitpunkt drei (dritte Klasse; n = 826).

Das Faktorenmodell mit den zwei Facetten Singen und Musizieren kann nicht gegen ein Faktorenmodell mit einem übergeordneten Faktor 2. Ordnung getestet werden, da es empirisch zu dem Faktorenmodell 1. Ordnung äquivalent ist (vgl. Geiser, 2010). Die Fitwerte weisen darauf hin, dass das theoretisch implizierte Faktorenmodell mit einem genesteten Generalfaktor gegenüber den anderen getesteten Modellen vorzuziehen ist. Die Struktur des postulierten Faktorenmodells für die beiden Messzeitpunkte ist in Abbildung 2 dargestellt.

Abb. 2: Prozentuale Häufigkeiten des Merkmals „Kind erhält derzeit Instrumentalunterricht“ nach Angaben der Eltern von der ersten bis zur vierten Klasse über die Gruppen.



Anmerkungen: A steht für affektive Komponenten, K für kognitive Komponenten des Fähigkeits-selbstkonzepts. Es werden die Faktorladungen für alle drei Messzeitpunkte angegeben: 1. MZP/2. MZP/3.MZP. Nicht signifikante Faktorlösungen sind mit n.s. gekennzeichnet. Alle anderen Faktorladungen sind signifikant (Musizieren  $p < .001$ ; Singen  $p < .05$ ;  $g < .05$ ). Die Korrelationen zwischen  $g$  und den Subskalen Musizieren und Singen wurden unterbunden. Die Korrelationen zwischen den Subskalen sind nicht signifikant. Die Modellfits können Tabelle 2 entnommen werden.

Die Fitwerte dieses Modells liegen bis auf den *RMSEA* für das Modell des zweiten und dritten Messzeitpunkts in den von Hu und Bentler (1999) berichteten Cut-Off-Grenzen (vgl. Tabelle 1). Nach Browne und Cudeck (1993) führen Werte des *RMSEA*, die größer als .08 sind, zu einer Ablehnung des Modells. Da der *RMSEA* für das vorliegende Modell nur geringfügig von den strengerer Richtlinien von Hu und Bentler (1999) abweicht, wird es angenommen.

Betrachtet man die Faktorladungen auf die latente Variable Singen, dann fällt auf, dass diese über die Zeit stark heterogen und instabil hinsichtlich ihrer Ladungsmuster sind. Sie variieren zum Teil zwischen .60 ( $p < .05$ ) und .18 (nicht signifikant) für das Item „Singen kann ich richtig gut“. Die Faktorladungen für das Item „Singen fällt mir leicht“ verfehlen über alle Messzeitpunkte hinweg das 5 Prozent Signifikanzniveau, laden jedoch hochsignifikant auf dem genesteten Generalfaktor  $g$  (.70 bis .86,  $p < .001$ ). Auch das Item „Singen kann ich richtig gut“ lädt auf dem genesteten Generalfaktor mit .68 bis .91 ( $p < .01$ ) moderat bis hoch. Die aufgeklärte Varianz für den latenten Faktor Singen beträgt für den ersten und dritten Messzeitpunkt .66 und .64 ( $p < .001$ ). Die aufgeklärte Varianz zum zweiten Messzeitpunkt ist nicht signifikant.

Die Facette Musizieren weist insgesamt stabilere Ladungsmuster auf. Die Höhe der Faktorladungen über die Zeit variiert nur mäßig. Zudem sind alle Ladungen auf dem 1 Prozent Niveau signifikant. Die höchsten Ladungen können für das Item „Musikmachen kann ich

richtig gut“ mit .69 ( $p < .001$ ) bis .83 ( $p < .001$ ) beobachtet werden. Die aufgeklärte Varianz beträgt .44 ( $p < .001$ ) bis .57 ( $p < .001$ ).

Für den genesteten Generalfaktor zeigen sich die Faktorladungen der Items, die zusätzlich auf die latente Variable Singen laden als instabil und sind zum Teil auch nicht signifikant. Die höchsten Ladungen auf den Generalfaktor haben die Items „Ich bin gut im Singen“, „Singen fällt mir leicht“ und „Singen kann ich richtig gut“. Insgesamt ist die Varianzaufklärung des genesteten latenten Faktors mit .22 bis .29 allerdings nicht besonders hoch.

## 4.2 Konvergente Validität

Für das genestete Faktorenmodell mit den Subfacetten Singen und Musizieren sowie dem genesteten Generalfaktor  $g$  wurde weiterhin überprüft, ob Zusammenhänge mit der Einschätzung der musikalischen Fähigkeiten durch die Lehrkraft und durch die Eltern beobachtet werden können (vgl. Tabelle 2).

Tab. 2: Zusammenhang des musikalischen Selbstkonzepts mit der Einschätzung der Fähigkeiten durch relevante Bezugspersonen zu drei Messzeitpunkten.

		Musikalität des Kindes aus Elternsicht	Freude am Singen aus Elternsicht	Musikalität des Kindes aus Lehrersicht
Generalfaktor $g$	MZP 1	0.07 n.s.	0.13 n.s.	0.23 *
	MZP 2	0.24 ***	0.35 **	0.24 ***
	MZP 3	0.12 n.s.	0.34 **	0.15 **
Musizieren	MZP 1	0.21 **	0.20 n.s.	0.13 n.s.
	MZP 2	0.14 *	-0.07 n.s.	0.18 ***
	MZP 3	0.36 ***	0.09 n.s.	0.28 ***
Singen	MZP 1	0.19 **	0.38 ***	0.40 **
	MZP 2	-0.03 n.s.	0.30 *	0.02 n.s.
	MZP 3	0.20 **	0.43 ***	0.10 n.s.

Anmerkungen: \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; nicht signifikante Parameter wurden mit *n.s.* gekennzeichnet. Die Korrelationen beruhen auf Korrelationen zwischen manifesten Skalenmittelwerten der beobachteten Außenkriterien mit den geschätzten latenten Subskalen aus dem genesteten Faktorenmodell. MZP1 = Messzeitpunkt 1 (erste Klasse), MZP2 = Messzeitpunkt zwei (zweite Klasse), MZP3 = Messzeitpunkt drei (dritte Klasse).

Am stärksten zeigt sich der Zusammenhang des Selbstkonzepts Singen mit der Einschätzung der Freude des Kindes am Singen durch die Eltern zu allen drei Messzeitpunkten (MZP1:  $r = .38, p < .001$ ; MZP2:  $r = .30, p < .05$ ; MZP3:  $r = .43, p < .001$ ). Dies ist erwartungskonform, da die Subskala Singen tatsächlich Aspekte der wahrgenommenen Freude (affektive Komponenten) erfasst.

In der vorliegenden Studie können bereits in der ersten Klasse substantielle Zusammenhänge beobachtet werden (Tab. 3). Die Korrelation für die Einschätzungen der Musikalität des Kindes aus Elternsicht und die Einschätzung musikalischer Fähigkeiten durch die Lehrkraft mit den Fähigkeitsselbsteinschätzungen der Kinder im Bereich Singen beträgt  $r = .19$  ( $p < .01$ ) und  $r = .40$  ( $p < .01$ ). Der moderate Zusammenhang für die Einschätzung musikalischer Fähigkeiten durch die Lehrkraft und die Fähigkeitsselbsteinschätzung im Singen ist lediglich für den ersten Messzeitpunkt zu beobachten und verliert in den zwei darauf folgenden Befragungen an Substanz. Die Freude am Singen aus Elternperspektive steht zu keinem Messzeitpunkt in einem signifikanten Zusammenhang mit der Fähigkeitsselbsteinschätzung im Musizieren. Der Zusammenhang der Einschätzung der musikalischen Fähigkeiten des Kindes durch die Lehrkraft und durch die Eltern mit der Fähigkeitsselbsteinschätzung bleibt über die zwei Messzeitpunkte relativ stabil. Insgesamt sprechen die Befunde mit entwicklungsbedingten Einschränkungen jedoch für die Annahme konvergenter Validität für die zwei Facetten Singen und Musizieren.

Tab. 3: Retest-Reliabilitäten: Korrelationen der Subskalen zu drei Messzeitpunkten (Korrelation nach Spearman).

	SK Singen MZP1	SK Singen MZP2
SK Singen MZP2	0.28 <i>n.s.</i>	-
SK Singen MZP3	0.26 <i>n.s.</i>	0.61***
	SK Musizieren MZP1	SK Musizieren MZP2
SK Musizieren MZP2	0.31**	-
SK Musizieren MZP3	0.21**	0.34***
	genesteter Generalfaktor <i>g</i> MZP1	genesteter Generalfaktor <i>g</i> MZP2
genesteter Generalfaktor <i>g</i> MZP2	0.41 <i>n.s.</i>	-
genesteter Generalfaktor <i>g</i> MZP3	0.19 <i>n.s.</i>	0.47***

*Anmerkungen:* MZP1 = Messzeitpunkt 1 (erste Klasse), MZP2 = Messzeitpunkt zwei (zweite Klasse), MZP3 = Messzeitpunkt drei (dritte Klasse). \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; nicht signifikante Parameter wurden mit *n.s.* gekennzeichnet.

### 4.3 Reliabilität

Hinweise zur Retest-Reliabilität liefern die Korrelationen der jeweiligen Subdimensionen über die Zeit, wobei die Befragungen in einem zeitlichen Abstand von etwa einem Jahr durchgeführt wurden.

Die Korrelationen der Subskala Singen sind für den zweiten und dritten Messzeitpunkt mit  $r = .61$  ( $p < .001$ ) vergleichsweise hoch und substantiell. Für die Selbsteinschätzung im Singen besteht kein Zusammenhang zwischen dem Urteil zum ersten und zweiten Messzeitpunkt. Die Zusammenhänge für die Subskala Musizieren liegen zwischen  $r = .21$  ( $p < .01$ ) für den zweiten Messzeitpunkt und  $r = .34$  ( $p < .001$ ) für den dritten Messzeitpunkt. Es handelt sich somit um moderate, jedoch relativ stabile Zusammenhänge über die Zeit. Die Retest-Reliabilität für den genesteten Generalfaktor ist lediglich für den zweiten und dritten Messzeitpunkt substantiell ( $r = .47$ ;  $p < .001$ ). Insgesamt zeigen sich hochsignifikante mittlere Zusammenhänge vor allem für die Korrelation zwischen zweitem und drittem Messzeitpunkt.

#### *Statistische Invarianz*

Als methodisches Vorgehen erfolgt die Invarianztestung im Rahmen von konfirmatorischen Faktorenanalysen im Strukturgleichungsansatz, wie sie bereits u. a. bei Reise, Widaman und Pugh (1993), Meredith (1993), Millsap und Yun-Tein (2004) sowie Weiber und Mühlhaus (2010) beschrieben wurde.

Im vorliegenden Fall wird statistische Invarianz zunächst für das genestete Faktorenmodell für die drei vorliegenden Messzeitpunkte nach den Vorgaben von Steenkamp und Baumgartner (1998) überprüft. Die Fitwerte für die betrachteten Modelle liegen innerhalb der von Hu und Bentler (1999) berichteten Cut-Off-Grenzen (vgl. Tabelle 4). Ist der für genestete Modelle entwickelte  $\Delta \chi^2$ -Test signifikant, dann bedeutet das, dass das restringierte Modell statistisch bedeutsam von dem unrestringierten Modell (Basismodell) abweicht. Für das Modell der schwachen (metrischen) Invarianz liegt keine statistisch bedeutsame Abweichung vom Basismodell vor, so dass das Vorliegen schwacher Invarianz angenommen werden kann. Das bedeutet, dass sich die Faktorladungen über die verschiedenen Messzeitpunkte nicht unterscheiden, jedoch die Schwellenwerte und die Residualvarianzen. Nach Meredith ist vollständige strikte Invarianz in der Realität kaum auffindbar, so dass in der Praxis häufig einzelne Parameter wieder freigegeben werden, so dass diese über die Zeit variieren können. Strenge partielle Messinvarianz kann im vorliegenden Fall nur über die Freisetzung nahezu aller Schwellenwerte sowie einer Auflockerung der Faktorenstruktur (durch das Zulassen von Nebenladungen von Items) erreicht werden. Da dieser Befund gegen ein stabiles

Konstrukt mit äquivalenten Schwellenwerten über die Zeit spricht, werden keine Freisetzen-  
gen vorgenommen.

Tab. 4: Überprüfung statistischer Invarianz über die Zeit: Modellfits der unrestringierten und restrin-  
gierten Modelle ( $n = 794$ ).

	$\chi^2$	df	$p$	$\Delta \chi^2$	$\Delta$ df	$\Delta p$	CFI	TLI	RMSEA
konfigural	461.21	325	.000				.995	.993	.023
schwach	463.00	359	.000	37.38	34	.317	.996	.995	.019
streng	601.85	416	.000	177.31	91	.000	.993	.992	.024
strikt	669.61	436	.000	246.02	111	.000	.991	.991	.026

In einem nächsten Schritt wurde die geschlechtsbezogene statistische Invarianz sowie in einem weiteren Schritt die Invarianz für Kinder, die ein Instrument spielen gegenüber Kin-  
dern, die kein Instrument spielen für jeden Messzeitpunkt separat überprüft. Es kann strenge  
(skalare) statistische Invarianz hinsichtlich des Geschlechts angenommen werden (vgl. Ta-  
belle 5).

Tab. 5: Überprüfung geschlechtsbezogener und gruppenspezifischer statistischer Invarianz. Modellfits der unrestringierten und restringierten Modelle.

Geschlecht ( <i>n</i> = 786)									
MZP 1	$\chi^2$	df	<i>p</i>	$\Delta \chi^2$	$\Delta$ df	$\Delta$ p	CFI	TLI	RMSEA
konfigural	86.54	55	.004				.996	.993	.038
schwach	91.32	67	.026	10.29	12	.590	.997	.996	.030
streng	109.47	82	.023	30.24	27	.304	.996	.996	.029
strikt	138.06	95	.003	62.63	40	.013	.994	.994	.034
MZP 2									
konfigural	111.28	55	.000				.994	.990	.050
schwach	115.00	66	.000	17.91	11	.084	.994	.992	.043
streng	132.87	82	.000	38.03	27	.077	.994	.994	.039
strikt	161.08	95	.000	69.81	40	.002	.993	.993	.041
MZP 3									
konfigural	152.83	55	.000				.994	.990	.066
schwach	127.00	67	.000	8.94	12	.708	.996	.995	.047
streng	144.48	82	.000	30.22	27	.305	.996	.996	.043
strikt	224.66	95	.000	106.22	40	.000	.992	.992	.058

Gruppe (mit Instrumentalunterricht/ ohne Instrumentalunterricht; $n = 624$ )									
MZP 1	$\chi^2$	df	$p$	$\Delta \chi^2$	$\Delta df$	$\Delta p$	CFI	TLI	RMSEA
MZP 1									
konfigural	63.62	55	.199				.999	.998	.022
schwach	74.68	66	.217	12.20	11	.349	.999	.998	.021
streng	90.05	82	.254	27.67	27	.428	.999	.999	.018
strikt	106.37	95	.200	45.11	40	.267	.998	.998	.020
MZP 2									
konfigural	99.55	55	.000				.995	.992	.056
schwach	103.47	66	.002	15.01	11	.182	.996	.994	.047
streng	113.59	82	.012	30.59	27	.288	.996	.996	.039
strikt	138.44	95	.002	60.20	40	.021	.995	.995	.042
MZP 3									
konfigural	112.08	55	.000				.996	.994	.064
schwach	105.15	67	.002	14.46	12	.272	.998	.997	.048
streng	152.90	82	.000	58.71	27	.000	.996	.995	.059
strikt	-	-	-	-	-	-	-	-	-

Anmerkungen: MZP1 = Messzeitpunkt 1 (erste Klasse), MZP2 = Messzeitpunkt zwei (zweite Klasse), MZP3 = Messzeitpunkt drei (dritte Klasse).

Die Invarianztestung über den Gruppierungsfaktor Instrumentalunterricht versus kein Instrumentalunterricht deutet auf fehlende Äquivalenz des Instruments zur Erfassung des musikalischen Selbstkonzepts mit den Facetten Singen und Musizieren hin (vgl. Tabelle 5). Zum ersten Messzeitpunkt liegt strikte statistische Invarianz vor, d. h. das Instrument misst über die Gruppen identisch und vollständig reliabel (Weiber & Mühlhaus, 2010). Für den zweiten Messzeitpunkt kann strenge (skalare) statistische Invarianz angenommen werden. Demnach sind Mittelwertvergleiche zwischen den Gruppen zulässig. Für den dritten Messzeitpunkt deutet der  $\Delta \chi^2$ -Test darauf hin, dass das restringierte Modell signifikant von dem Basismodell abweicht. Hier kann lediglich schwache (metrische) statistische Invarianz angenommen werden. Somit nimmt die Messäquivalenz über die drei Messzeitpunkte im Hinblick auf einen Gruppenvergleich von Kindern, die Instrumentalunterricht erhalten gegenüber Kindern, die keinen Instrumentalunterricht erhalten ab.

#### **4.4 Geschlechts- und Gruppenunterschiede**

Da aufgrund der vorangegangenen Analysen zur statistischen Invarianz über die Zeit das Vorliegen strenger oder skalarer Invarianz in Frage gestellt werden muss, können Mittelwertvergleiche des hier vorgestellten Instruments über die Zeit nicht vorgenommen werden. Zulässig ist lediglich die Überprüfung von Beziehungen der Konstrukte im Strukturgleichungsansatz (Weiber & Mühlhaus, 2010), was aufgrund der Fülle der hier dargestellten Informationen an anderer Stelle erfolgen soll. Für das musikalische Selbstkonzept mit den Facetten Singen und Musizieren sowie dem genesteten Generalfaktor liegt jedoch im Hinblick auf den Gruppierungsfaktor Geschlecht strenge statistische Invarianz vor, so dass es zulässig ist, die im Modell geschätzten Mittelwerte von Jungen und Mädchen miteinander zu vergleichen. In Tabelle 6 sind die auf Grundlage der jeweils gültigen Invarianzmodelle geschätzten Mittelwerte, Standardabweichungen und  $p$ -Werte angegeben. Für alle drei Messzeitpunkte kann beobachtet werden, dass Jungen im Singen tendenziell ein niedrigeres Selbstkonzept aufweisen als Mädchen. Dieser Befund ist jedoch lediglich für den dritten Messzeitpunkt statistisch bedeutsam ( $M = -3.57$ ,  $SD = 0.59$ ,  $p < .001$ ). Ähnliches gilt für das Selbstkonzept Musizieren. Hier zeigt sich tendenziell eine positivere Einschätzung der Jungen gegenüber den Mädchen. Allerdings ist dieser Unterschied für die ersten beiden Messzeitpunkte nicht signifikant. Statistisch relevant ist der Unterschied im Selbstkonzept Musizieren lediglich zu Messzeitpunkt drei, hier jedoch zugunsten der Mädchen ( $M = -0.69$ ,  $SD = 0.18$ ,  $p < .01$ ). Diese Interpretation ist aufgrund von Verstößen gegen die statistische Messinvarianz nicht zulässig.

Tab. 6: Modellbasierte geschätzte Mittelwertunterschiede zwischen Jungen und Mädchen (Referenzvariable: Mädchen) und zwischen Kindern mit und ohne Instrumentalunterricht (Referenzvariable: ohne Instrumentalunterricht) zu drei Messzeitpunkten.

	Jungen			Kinder mit Instrumentalunterricht		
	M	SD	<i>p</i>	M	SD	<i>p</i>
<i>MZP 1</i>						
Singen	-0.21	0.40	.602	0.07	0.25	.769
Musizieren	0.16	0.21	.445	0.56	0.17	.001
<i>g</i> Generalfaktor	-0.73	0.15	.000	0.10	0.13	.429
<i>MZP 2</i>						
Singen	-1.10	0.69	.112	-0.45	0.42	.282
Musizieren	0.06	0.29	.851	0.35	0.21	.103
<i>g</i> Generalfaktor	-0.78	0.30	.009	-0.03	0.21	.899
<i>MZP 3</i>						
Singen	-3.57	0.59	.000	1.99	1.04	.057
Musizieren	-0.69	0.26	.008	2.46	0.47	.000
<i>g</i> Generalfaktor	0.06	0.18	.721	-0.71	0.32	.028

Anmerkungen: MZP1 = Messzeitpunkt 1 (erste Klasse), MZP2 = Messzeitpunkt zwei (zweite Klasse), MZP3 = Messzeitpunkt drei (dritte Klasse).

Die Mittelwertvergleiche für den genesteten Generalfaktor unterscheiden sich ebenfalls zwischen den ersten beiden und dem letzten Messzeitpunkt. Zunächst unterscheiden sich Mädchen und Jungen hier signifikant hinsichtlich des Generalfaktors für die ersten beiden Messzeitpunkte. Jungen haben ein geringeres allgemeines musikalisches Selbstkonzept als Mädchen. Zum dritten Messzeitpunkt verliert dieser Unterschied statistisch an Bedeutsamkeit.

Von besonderem Interesse ist zudem wie sich das Selbstkonzept von Kindern mit und ohne Instrumentalunterricht entwickelt. Da für die ersten beiden Messzeitpunkte mindestens strenge statistische Invarianz vorliegt, sind Mittelwertvergleiche sinnvoll zu interpretieren.

Der Vollständigkeit halber werden auch die Werte für den dritten Messzeitpunkt angegeben. Diese gilt es jedoch, aufgrund fehlender strenger Invarianz, mit Vorsicht zu interpretieren. In Tabelle 6 sind die geschätzten Mittelwerte, Standardabweichungen und  $p$ -Werte aus dem jeweils gültigen Invarianzmodell aufgeführt. Für den ersten Messzeitpunkt kann beobachtet werden, dass Kinder, die Unterricht auf einem Musikinstrument erhalten, ihre Fähigkeiten im Musizieren signifikant besser bewerten als Kinder, die keinen Unterricht erhalten ( $M = -0.56$ ,  $SD = 0.17$ ,  $p < .001$ ). Die Mittelwerte für Singen und den genesteten Generalfaktor sind geringfügig höher für Kinder mit Instrumentalunterricht, aber nicht statistisch bedeutsam. Zum zweiten Messzeitpunkt ändern sich für Singen und den Generalfaktor die Vorzeichen. Allerdings ist keiner der im Modell geschätzten Mittelwerte signifikant. Mit großer Vorsicht kann für den dritten Messzeitpunkt die Tendenz beobachtet werden, dass sich Kinder, die Instrumentalunterricht erhalten, hinsichtlich ihrer Fähigkeiten im Musizieren positiver einschätzen als Kinder, die keinen Unterricht erhalten. Die Einschätzung für das Fähigkeits-selbstkonzept Singen deutet in dieselbe Richtung ( $M = 1.99$ ,  $SD = 1.04$ ,  $p < .1$ ). Die Einschätzung des genesteten Faktors ist für die Kinder mit Unterricht jedoch negativ ( $M = -0.71$ ,  $SD = 0.32$ ,  $p < .01$ ).

## 5. Diskussion

Das Ziel der Arbeit bestand darin, ein standardisiertes Instrument zur Erfassung des musikalischen Selbstkonzepts im Grundschulalter zu entwickeln und erste Hinweise auf die Reliabilität und Validität des Instruments in quer- und längsschnittlichen Betrachtungen zu finden.

Anhand von konfirmatorischen Faktorenanalysen konnte gezeigt werden, dass sich die Items eindeutig den zwei Bereichen des Singens und Musizierens zuordnen lassen. Ein weiterer genesteter Generalfaktor erklärt zusätzlich itemspezifische Varianz der zwei latenten Selbstkonzeptfacetten Singen und Musizieren. Modellvergleiche zeigten, dass das genestete Faktorenmodell die vorliegenden Daten besser beschreibt als ein Generalfaktorenmodell oder ein Faktorenmodell zweiter Ordnung. Das könnte ein Hinweis darauf sein, dass die Entwicklung des Selbstkonzepts von Kindern in der Grundschule noch nicht abgeschlossen ist. Nach Harter (1982) und Harter und Pike (1984) nimmt die Differenzierungsfähigkeit, zwischen einzelnen Kompetenzbereichen unterscheiden zu können, im Laufe der Grundschulzeit sowie mit fortschreitender kognitiver Entwicklung der Kinder zu. Diese Entwicklung wird auch für die Verarbeitung von Rückmeldungen durch wichtige Bezugspersonen sowie die Einbindung solcher Informationen in das eigene Selbstkonzept berichtet. In der vorliegenden Studie können bereits substantielle Zusammenhänge zwischen dem Selbstkonzept Singen bzw. Musizieren und dem Lehrerurteil sowie dem Elternurteil für den ersten Messzeitpunkt berichtet werden. Inwiefern die Rückmeldungen der Bezugspersonen bereits für die Ausprägung der Fähigkeits-selbstkonzepte verantwortlich sind, kann anhand der vorliegenden Daten nicht geklärt werden.

Anhand der konfirmatorischen Faktorenanalysen zeigte sich, dass die Faktorladungen für die Subskala Singen vergleichsweise instabil sind und zum Teil nicht signifikante Ladungen

aufweisen (vgl. Abbildung 1). Trotz substantieller Zusammenhänge mit der Einschätzung relevanter Bezugspersonen kann daher angenommen werden, dass die Fähigkeitsselbsteinschätzung für den Bereich Singen für Kinder noch sehr abstrakt ist. Betrachtet man die Items, die auf den latenten Faktor Singen laden, dann fällt auf, dass die Items anders als bei der Subskala Musizieren inhaltlich auch den Aspekt der erlebten Freude beim Singen umfassen. Die Verwendung von affektiven Komponenten des Selbstkonzepts sind umstritten und werden u. a. bei Schwanzer (2002) und Schwanzer et al. (2005) diskutiert. So argumentiert Marsh (1986), dass sowohl kognitiv-evaluative Elemente als auch affektive Elemente in die Ausprägung des Selbstkonzepts einfließen. Dabei gibt es starke Zusammenhänge der affektiven Komponenten mit der empfundenen Selbstnähe und dem Interessenskonzept (Kessels & Hannover, 2004). Die Erfassung des Selbstkonzepts erfolgte im schulischen, also einem mit Leistung assoziierten Kontext. Bei dem Faktor Singen fehlt es den Kindern vermutlich an objektiven Vergleichskriterien, da die eigenen Fähigkeiten im Singen nur schwer mit Klassenkameraden verglichen werden können. Zudem wird der Gesang vergleichsweise wenig von den Lehrkräften geschult oder Fehler beispielsweise beim Halten der richtigen Tonlage etc. zurückgemeldet, wie es für das Spielen eines Musikinstruments vermutlich eher der Fall ist. Dafür sprechen auch die geringen Zusammenhänge zwischen der Beurteilung der musikalischen Fähigkeiten durch die Lehrkraft und den Selbsteinschätzungen der eigenen Fähigkeiten im Singen zum dritten und vierten Messzeitpunkt. Dementsprechend scheint für Kinder die inhaltliche Komponente der Freude am Singen ein (vorerst) wichtiger Bestandteil der Fähigkeitsselbsteinschätzung zu sein. Interessant ist in diesem Zusammenhang, dass die affektiven Items zum ersten Messzeitpunkt nicht auf den Generalfaktor laden, in den folgenden Messzeitpunkten jedoch schon. Das könnte bedeuten, je länger die Kinder musizieren, desto mehr gleichen sich affektive und kognitive Komponenten aneinander an.

Insgesamt zeigen sich hochsignifikante mittlere Korrelationen für die einzelnen Facetten zwischen dem zweiten und dritten Messzeitpunkt. Helmke (1991) beobachtete, dass die zeitliche Stabilität des Selbstkonzepts in den ersten Grundschuljahren noch relativ gering ist. Er fand Korrelationen für das schulische Fähigkeitsselbstbild zwischen der ersten Klasse und der zweiten Klasse von  $r = .24$  ( $p < .01$ ). Die zum Konstrukt des allgemeinen Fähigkeitsselbstkonzepts bei Helmke äquivalente Entwicklung der Subskala Singen und des allgemeinen Generalfaktors kann als Hinweis auf die Validität des vorliegenden Instruments interpretiert werden.

Das vorgestellte Instrument zur Erfassung musikalischer Fähigkeiten in den Bereichen Singen und Musizieren wurde im Rahmen von Multigruppenvergleichen im Strukturgleichungsansatz auf zeitliche, geschlechtsbezogene und gruppenspezifische (Instrumentalunterricht versus kein Instrumentalunterricht) Invarianz überprüft.

Das Instrument weißt zu allen Messzeitpunkten gute Anpassungswerte auf. Dennoch liegen Hinweise auf den Verstoß von statistischer Invarianz über die Zeit vor. Es kann lediglich schwache Invarianz bestätigt werden. Dies bedeutet, dass eine Entwicklung des musikalischen Selbstkonzepts aufgrund von unzulässigen Parametervariationen nicht abgebildet wer-

den kann. Das Instrument zur Erfassung des musikalischen Selbstkonzepts kann im Hinblick auf Geschlechtsunterschiede jedoch als hinreichend äquivalent gelten, da ein Minimum an strenger Invarianz vorliegt. Mittelwertvergleiche deuten darauf hin, dass sich Mädchen und Jungen in der ersten und zweiten Klasse nicht hinsichtlich ihres Fähigkeitsselbstkonzepts im Musizieren voneinander unterscheiden. Dies entspricht auch dem Befund von Kranefeld und Busch (2012) die in der zweiten Klasse keinen Unterschied zwischen Mädchen und Jungen beobachten konnten. In der dritten Klasse jedoch unterscheiden sich Mädchen und Jungen in der vorliegenden Stichprobe signifikant hinsichtlich ihres Fähigkeitsselbstkonzepts Musizieren. Jungen weisen diesbezüglich ein um  $-0.69$  ( $SD = 0.26$ ;  $p < .01$ ) Skalenwertpunkte geringeres Selbstkonzept auf als Mädchen. Für die Gruppierungsvariable Kinder mit und ohne Instrumentalunterricht zeigen sich über die Zeit zunehmend Verletzungen der Invarianzannahme. Zum dritten Messzeitpunkt kann lediglich metrische Invarianz angenommen werden. Bezüglich fehlender strenger Invarianz über die Zeit und hinsichtlich der Gruppierungsvariable „Instrumentalunterricht“ scheint es möglich, dass die kognitive Entwicklung der Kinder über die Zeit (Harter & Pike, 1984) sowie Effekte des *Internal-External-Frame of Reference-Models* (Marsh, 1986) für die fehlende Stabilität verantwortlich sind. Wie bereits eingangs erwähnt, wird die Ausbildung des Fähigkeitsselbstkonzepts maßgeblich von bedeutenden Bezugspersonen, sogenannten „significant others“ geformt. Bezugsgruppeneffekte, wie sie beispielsweise in Schulklassen, in denen Instrumentalunterricht angeboten wird, zu erwarten sind, können an dieser Stelle nicht weiter überprüft werden, sind jedoch von hohem Interesse und vermutlich hochgradig relevant für die Selbstkonzeptgenese im musikalischen Bereich (Fritzsche et al., 2011). Hierauf deuten auch die Mittelwertvergleiche der latenten Variablen hin. Unterschiede zwischen Kindern, die Instrumentalunterricht erhalten gegenüber Kindern, die keinen erhalten, verlieren am Ende der zweiten Klasse zunächst an statistischer Bedeutsamkeit. Das JeKi-Konzept in Nordrhein-Westfalen sieht für die zweite Klasse Unterricht auf einem Musikinstrument nach eigener Wahl vor, nachdem diese in der ersten Klasse vorgestellt wurden. In Hamburg beginnt der JeKi-Unterricht in der zweiten Klasse wobei die Vorstellung der Instrumente im Klassenverbund im Vordergrund steht. Es kann angenommen werden, dass soziale Vergleichsprozesse im Musizieren und Singen für diese Ergebnisse verantwortlich sind. Möglicherweise führen Vergleiche mit Mitschülern zu einer Anpassung der Wahrnehmung der eigenen Fähigkeiten in Relation zum Fähigkeitsniveau der Lerngruppe. Des Weiteren können zudem Vergleiche von Leistungen in anderen Fächern zu einer Relativierung der Wahrnehmung der eigenen musikalischen Fähigkeiten führen. Das Vorliegen von sozialen und dimensional Vergleichen kann anhand des vorliegenden Datensatzes nicht überprüft werden, da die Fähigkeiten bzw. Leistungen der Kinder in Musik und in den anderen Schulfächern nicht erfasst wurden. Allerdings sind solche reziproken Effekte in der Literatur gut dokumentiert (u. a. Marsh, 1984; Eccles et al., 1993; Fritzsche et al., 2011; Helmke, 1991).

## 6. Limitationen und Implikationen für zukünftige Arbeiten

Es muss kritisch angemerkt werden, dass durch die inhaltliche Beschränkung auf die beiden Aspekte Singen und Musizieren nur ein Ausschnitt des musikalischen Selbstkonzepts erfasst wird. Bereiche, die nicht betrachtet werden, betreffen z. B. die Aspekte des Musikhörens, das Rezipieren oder beispielsweise die Notenkunde.<sup>3</sup> Dies ist möglicherweise eine Erklärung für die in dieser Studie beobachtete Modellstruktur der zwei Subfacetten mit einem genesteten Generalfaktor. So kann angenommen werden, dass der genestete Generalfaktor itemspezifische Varianz beinhaltet, die nicht über die zwei Facetten aufgeklärt wird. Eine weitere Facette, wie das Selbstkonzept des Musikhörens kann vermutlich nur mit großen Einschränkungen altersgemäß und inhaltlich valide erfasst werden, so dass auf die Formulierung von Items zur Erfassung dieser Facette im Vorwege verzichtet wurde. Auch weitere mit musikalischen Fähigkeiten assoziierte Bereiche wurden bei der Entwicklung des Instruments nicht berücksichtigt, da davon auszugehen ist, dass Kinder eine dementsprechende Differenzierung in jungen Jahren noch nicht vornehmen können, bzw. ihnen die Erfahrung für bestimmte Bereiche fehlt.

Zudem wurde das Instrument mit dem Fokus auf die Begleitforschung des Programms JeKi entwickelt. Im Kontext der Programmevaluation sollte ein Einblick in die Selbstkonzeptgenese unter dem Einfluss von Instrumentalunterricht an Grundschulen vorgenommen werden. Eine Weiterentwicklung des vorgestellten Instruments mit weiteren Subfacetten wäre wünschenswert. Des Weiteren haben statistische Kriterien zur Annahme des genesteten Faktorenmodells geführt. Dieses Modell ist schwer zu interpretieren, da der Generalfaktor zusätzliche itemspezifische Varianz aufklärt, die jedoch kaum inhaltlich gefasst werden kann. Dementsprechend sollte für zukünftige Analysen möglicherweise ein theoretisch abgeleitetes Faktorenmodell mit zwei Facetten Singen und Musizieren genutzt werden. Der Mehrwert des genesteten Modells liegt hier lediglich auf einer besseren Modellanpassung des postulierten Modells auf die empirischen Daten.

Eine Schwäche des vorgestellten Instruments zur Erfassung des Fähigkeitsselbstkonzepts Musizieren kann darin ausgemacht werden, dass auch Singen theoretisch zur Aktivität des Musizierens bzw. Musikmachens gehört. Es kann jedoch davon ausgegangen werden, dass Kinder im Grundschulalter noch nicht über die kognitiven Fähigkeiten zur Abstraktion dieses Sachverhalts verfügen, so dass das vorgestellte Instrument für die Adressatengruppe dieser Studie uneingeschränkt verwendet werden kann.

---

<sup>3</sup> Leider existieren bisher nur wenige theoretische Vorarbeiten, die mögliche Differenzierungen von Selbstkonzeptfacetten von Kindern in den Blick nehmen. Die Arbeiten im Kontext des Projekts KoMus von Niessen, Lehmann-Wermser, Knigge und Lehmann (2008) sowie Knigge (2010) und Jordan, Knigge, Lehmann, Niessen & Lehmann-Wermser (2012) ermitteln theoretisch begründete Kompetenzbereiche, die in zukünftigen Forschungsarbeiten als Grundlage für die Entwicklung von verschiedenen Facetten des musikalischen Selbstkonzepts dienen können.

Weiterhin ist für die Selbstkonzepte Singen und Musizieren kritisch anzumerken, dass die Items „Ich bin gut im Singen“ bzw. „Ich bin gut im Musikmachen“ und „Singen kann ich richtig gut“ bzw. „Musikmachen kann ich richtig gut“ inhaltlich eine sehr starke Überschneidung aufweisen. Dennoch zeigte sich in den konfirmatorischen Faktorenanalysen, dass Kinder diesbezüglich eine inhaltliche Unterscheidung treffen. Dabei ist anzunehmen, dass die Items „Singen kann ich richtig gut“ und „Musikmachen kann ich richtig gut“ möglicherweise mit besonders guten Fähigkeiten assoziiert werden.

Zusammenfassend bestätigt die vorliegende Arbeit die Annahme, wonach Kinder bereits in den ersten Grundschuljahren mit Hilfe standardisierter Verfahren hinsichtlich ihrer musikalischen Fähigkeitsselbsteinschätzung befragt werden können. Analysen zur Messinvarianz deuten jedoch darauf hin, dass die Bedeutung einzelner Items sowie die Interpretation der Antwortkategorien über die Zeit und im Verlauf der kognitiven Entwicklung der Kinder variieren. Somit kann das vorliegende Instrument querschnittlich genutzt valide und reliable Informationen für Gruppenvergleiche vor allem über die ersten beiden Grundschuljahre liefern. Des Weiteren eignet es sich, um Strukturen musikalischer Fähigkeiten in den Bereichen des Singens und Musizierens abzubilden. Die Entwicklung der musikalischen Fähigkeiten über die Zeit kann jedoch nicht abgebildet werden. Es kann jedoch auch angenommen werden, dass die mangelnde Invarianz über die Zeit vor allem zu Lasten der noch andauernden kognitiven Entwicklung der Kinder sowie der Persönlichkeitsentwicklung geht. Dieses Problem gilt demnach nicht nur für die Entwicklung des vorliegenden Instruments, sondern trifft auch auf die Selbstkonzeptentwicklung in anderen Bereichen zu (Nonte, 2012). Der Umgang mit entwicklungsbedingten Veränderungen eines standardisierten Instruments über die Zeit ist bisher kaum in den Blick genommen worden und sollte in zukünftigen Studien differenzierter betrachtet werden.

## Literatur

- Benson, J. & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement*, 22, 231–240.
- Bernecker, C., Haag, L. & Pfeiffer, W. (2006). Musikalisches Selbstkonzept: Eine empirische Untersuchung. *Diskussion Musikpädagogik*, 29, 53–57.
- Borgers, N., de Leeuw, E. & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Méthodologie Sociologique*, 66, 60–75.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park: Sage.
- Busch, T. & Kranefeld, U. (2012). Sind Mädchen die besseren Optimisten im Umgang mit Musik? Prädiktoren des musikalischen Selbstkonzepts in der Klassenstufe 2. *Beiträge empirischer Musikpädagogik*, Bd. 3. Zugriff am 16.11.2012. Verfügbar unter [http://www.b-em.info/index.php?journal=ojs&page=article&op=view&path\[\]=70&path\[\]=200](http://www.b-em.info/index.php?journal=ojs&page=article&op=view&path[]=70&path[]=200).
- Busch, V. & Wagschal, J. (2010). Skalendokumentation zum SIGrun-Teilprojekt Präferenz. Unveröffentlichtes Manuskript.
- Eccles, J., Wigfield, A., Harold, R. D. & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, 64, 830–847.
- Fritzsche, E. S., Kröner, S. & Pfeiffer, W. (2011). Chorknaben und andere Gymnasiasten - Determinanten musikalischer Aktivitäten an Gymnasien mit unterschiedlichen Schulprofilen. *Journal for Educational Research Online*, 3, 94–118.
- Geiser, C. (2010). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Harter, S. (1982). The perceived competence scale for children. *Child Development*, 53, 87–97.
- Harter, S. & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development*, 55, 1969–1982.
- Harter, S. (1988). *Manual for the self-perception profile for adolescents*. Denver: University Press.
- Helmke, A. (1991). Entwicklung des Fähigkeitsselbstbildes vom Kindergarten bis zur dritten Klasse. In R. Pekrun & H. Fend (Hrsg.), *Schule und Persönlichkeitsentwicklung. Ein Resümee der Längsschnittforschung* (Der Mensch als soziales und personales Wesen, Bd. 11, S. 83–99). Stuttgart: Ferdinand Enke.

- Hu, L.-T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- 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*, 4, 500–521.
- Kessels, U. & Hannover, B. (2004). Empfundene "Selbstnähe" als Mediator zwischen Fähigkeitsselbstkonzept und Leistungskurswahlintentionen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36, 130–138.
- Kline, P. (1997). *An easy guide to factor analysis*. London: Routledge.
- Knigge, J. (2012). *Modellbasierte Entwicklung und Analyse von Testaufgaben zur Erfassung der Kompetenz „Musik wahrnehmen und kontextualisieren“*. Dissertationsschrift an der Universität Bremen. Zugriff am 06.03.2013. Verfügbar unter <http://elib.suub.uni-bremen.de/diss/docs/00012006.pdf>.
- Kröner, S., Schwanzer, A. D. & Dickhäuser, O. (2009). Jenseits von Mozart – eine Pilotstudie zu Determinanten musikalischer Aktivitäten während der Grundschulzeit. *Psychologie in Erziehung und Unterricht*, 56, 224–238.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau*, 58, 103–117.
- Marsh, H. W., Barnes, J., Cairns, L., Tidman, M. (1984). Self-Description Questionnaire: Age and sex effects in the structure and level of self-concept for preadolescent children. *Journal of Educational Psychology*, 76, 940–956.
- Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, 22, 37–49.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129–149.
- Marsh, H. W. (1987). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement*, 24, 17–39.
- Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, 82, 623–636.
- Marsh, H. W. (1992). *Self Description Questionnaire (SDQ) III: A theoretical and empirical basis for the measurement of multiple dimensions of late adolescent selfconcept: An interim test manual and a research monograph*. Macarthur, New South Wales, Australia.

- Marsh, H. W., Byrne, B. M. & Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, 80, 366–380.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543.
- Millsap, R. E. & Yun-Tein, J. (2004). Assessing factorial invariance in ordered categorical measures. *Multivariate Behavioral Research*, 39, 479–515.
- Muthén, B. O. & Muthén, L. K. (2010). Mplus (Version 6.1) [Computer software].
- 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?*, 3–33. <http://www.zfkm.org/sonder08-niessenetal.pdf> [06.03.2013].
- Nonte, S. (2012). *Die Überprüfung von geschlechtsbezogener Messinvarianz des Fähigkeits-selbstkonzepts von Grundschulern in der Schuleingangsphase*. *Empirische Pädagogik*, 26(4), 478 – 503.
- Nonte, S. & Schwippert, K. (2012). Musikalische und sportliche Profile an Grundschulen – Auswirkungen auf Klassenklima und Selbstkonzept. *Beiträge empirischer Musikpädagogik*, 3, 2–25.
- Pfeiffer, W. (2007). Das musikalische Selbstkonzept: Effekte und Wirkungen. *Diskussion Musikpädagogik*, 33, 40–44.
- Rauer, W. & Schuck, K. D. (2004). *FEESS 1–2: Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern erster und zweiter Klassen*. Göttingen: Beltz Test GmbH.
- Reise, S. P., Widaman, K. F. & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552–566.
- Schwanzer, A. (2002). *Entwicklung und Validierung eines deutschsprachigen Instruments zur Erfassung des Selbstkonzepts junger Erwachsener* (Materialien aus der Bildungsforschung, Bd. 74). Berlin: Max-Planck-Institut für Bildungsforschung.
- Schwanzer, A. D., Trautwein, U., Lüdtke, O. & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica*, 51, 183–194.
- Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Self-Concept: Validation of construct interpretations. *Review of Educational Research*, 46, 407–441.

- Spychiger, M. (2007). „Nein, ich bin ja unbegabt und liebe Musik“. Ausführungen zu einer mehrdimensionalen Anlage des musikalischen Selbstkonzepts. *Diskussion Musikpädagogik*, 33, 9–20.
- Spychiger, M., Gruber, L. & Olbertz, F. (o. J.). *Musical self-concept: Presentation of a multi-dimensional model and its empirical analysis*. Verfügbar unter: [http://www.hfmdk-frankfurt.info/fileadmin/Dateien/Forschung\\_und\\_Projekte/MusicalSelfconcept\\_PaperESCOM09-\\_Spychiger\\_01.pdf](http://www.hfmdk-frankfurt.info/fileadmin/Dateien/Forschung_und_Projekte/MusicalSelfconcept_PaperESCOM09-_Spychiger_01.pdf) [19.04.2011].
- Steenkamp, J.-B. E. M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78–90.
- Vispoel, W. P. (1993). The development and validation of the arts self-perception inventory for adolescents. *Educational and Psychological Measurement*, 53, 1023–1033.
- Vispoel, W. P. (1995). Self-Concept in artistic domains: An extension of the Shavelson, Hubner, and Stanton (1976) Model. *Journal of Educational Psychology*, 87, 134–153.
- Vispoel, W. P. (1996). The development and validation of the arts self-perception inventory for adults. *Educational and Psychological Measurement*, 56, 719–735.
- Weiber, R. & Mühlhaus, D. (2010). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Berlin: Springer.

## **Autorin:**

### **Sonja Nonte**

Georg-August-Universität Göttingen  
Institut für Erziehungswissenschaft  
Waldweg 26  
37073 Göttingen

Email: [snonte@uni-goettingen.de](mailto:snonte@uni-goettingen.de)