

## Mixed-Methods-Design als Grundlage für die Entwicklung von Aufgaben zur Erfassung von Modellkompetenz von Grundschulkindern

Florian Böschl, Sarah Gogolin, Kim Lange-Schubert, Andreas Hartinger

### Angaben zur Veröffentlichung / Publication details:

Böschl, Florian, Sarah Gogolin, Kim Lange-Schubert, and Andreas Hartinger. 2019. "Mixed-Methods-Design als Grundlage für die Entwicklung von Aufgaben zur Erfassung von Modellkompetenz von Grundschulkindern." In *Methodologien der Forschungen zur Didaktik des Sachunterrichts*, edited by Hartmut Giest, Eva Gläser, and Andreas Hartinger, 115–37. Bad Heilbrunn: Julius Klinkhardt.



*Florian Böschl, Sarah Gogolin, Kim Lange-Schubert  
und Andreas Hartinger*

## **Mixed-Methods-Design als Grundlage für die Entwicklung von Aufgaben zur Erfassung von Modellkompetenz von Grundschulkindern**

### **1 Erfassung von Kompetenzen als methodische Herausforderung für die Forschung zum Sachunterricht**

Die empirische Forschung zur Didaktik des Sachunterrichts ist darauf angewiesen, valide interpretierbare Instrumente für die Erhebung abhängiger Variablen (z.B. bezüglich des Lernzuwachses von Schüler/innen) zu entwickeln. Dies ist insbesondere dann eine große Herausforderung, wenn Verfahren entwickelt werden sollen, mit denen – gemäß der Ziele des Sachunterrichts – Kompetenzen von Lernenden erhoben werden (AERA, APA & NCME 2014). Für den Sachunterricht ist Kompetenz im Perspektivrahmen als ein Zusammenspiel von Denk-, Arbeits- und Handlungsweisen und Inhaltsbereichen definiert (GDSU 2013). Versteht man Sachunterricht als vielperspektivisches Fach, ergibt sich die zusätzliche Schwierigkeit, dass Kompetenzen zu erheben sind, die über einen Inhaltsbereich oder eine Perspektive hinausgehen (vgl. Willard & Roseman 2010). Zudem sind die in der sachunterrichtsdidaktischen Forschung fokussierten Schüler/innen i.d.R. im Alter von sechs bis zehn Jahren, sodass die Konstruktion von Instrumenten die entwicklungsgemäßen Fähigkeiten und Fertigkeiten dieser Altersgruppe berücksichtigen muss, beispielsweise mit Blick auf die Lesefähigkeiten oder die Kapazitäten des Arbeitsgedächtnisses. Für Studien zu den Wirkungen bestimmter Maßnahmen, die ein hypothesenvergleichendes Design erfordern, ist es zudem erforderlich, Instrumente zur Hand zu haben, mithilfe derer ökonomisch auch größere Stichproben untersucht werden können (vgl. Baur 2015, Fleischer et al. 2013, Hartig, Klieme, & Leutner 2008).

In diesem Beitrag sollen am Beispiel der Modellkompetenz diese Herausforderung sowie unser Ansatz zur Bewältigung dieser Herausforderung aufgezeigt und diskutiert werden. Der Schwerpunkt des vorliegenden Beitrags ist

die Darstellung der Entwicklung von Testinstrumenten in einem Mixed-Methods-Design mit qualitativen und quantitativen Anteilen (s. Kap. 5). Charakteristisch für unsere Arbeit ist dabei die Verortung in bestehenden theoretischen Modellen (der Modellkompetenz) und die stete empirische Überprüfung der entwickelten Instrumente. Durch die Rückkopplung von theoretischen Annahmen mit den tatsächlich im Feld vorzufindenden Ausprägungen der Modellkompetenz ist es möglich, zu präziseren theoretischen Beschreibungen und validen Messwertinterpretationen zu kommen.

Dazu werden im Folgenden zunächst die die Untersuchung leitenden theoretischen Annahmen sowie der aktuelle Forschungsstand zur Modellkompetenz mit Blick auf den Sachunterricht skizziert. Im Anschluss an die daraus abgeleiteten, spezifischen Fragestellungen werden – mit Blick auf den Schwerpunkt dieses Forschungsbands – insbesondere die der Arbeit zugrunde gelegten erkenntnistheoretisch-methodologischen Überlegungen und deren methodentheoretisch begründete Umsetzung in Datengewinnung, -erhebung und -analyse ausführlicher beleuchtet. Daraus resultierende Ergebnisse zur Modellkompetenz im Sachunterricht der Grundschule werden abschließend diskutiert.

## 2 Theoretische Grundlagen zur Modellkompetenz

Die Fähigkeit, kompetent mit Modellen umgehen zu können, ist eine perspektivenübergreifende Denk-, Arbeits- und Handlungsweise. Sie wird insbesondere im naturwissenschaftlich-technischen Bereich national wie international explizit in den Zielen (sach-)unterrichtlichen Lernens aufgeführt (GDSU 2013, NGSS Lead States 2013, NRC 2011) – ist jedoch auch für die anderen Perspektiven des Sachunterrichts von Bedeutung, da in allen fachwissenschaftlichen Referenzdisziplinen mit Modellen gearbeitet wird. Dem kompetenten Umgang mit Modellen wird nicht zuletzt deshalb eine große Bedeutung zugewiesen, da man davon ausgeht, dass durch die Arbeit mit und das Denken in Modellen der Erwerb von konzeptuellem Wissen über die im Unterricht behandelten Themenbereiche gestärkt wird (vgl. Clement 2000). Durch die explizite Auseinandersetzung mit bestehenden Vorstellungen und Konzepten in Modellierungsprozessen sollen zudem Umstrukturierungsprozesse im Sinne des Lernens als „Conceptual Change“ (vgl. Strike & Posner 1992, Möller 2015, Lange & Hartinger 2014) sowie ein vertieftes Verständnis über das Wesen der Naturwissenschaften (vgl. Vo, Forbes, Zangori & Schwarz 2015) unterstützt werden.

Mit dem Begriff einer übergeordneten Modellkompetenz werden zwei Aspekte verknüpft: Dies sind zum einen (prozedurale) Fähigkeiten – also das

Modellieren (die Konstruktion, Anwendung, Evaluation und Überarbeitung von Modellen) und zum anderen das (deklarative) Modellverstehen (Wissen über die Tätigkeit des Modellierens als Erkenntnisprozess sowie das Verstehen von Modellen als Wissensformen) (vgl. Gilbert & Justi 2016, Nicolaou & Constantinou 2014, Schwarz et al. 2009). Das von Upmeier zu Belzen und Krüger (2010) entwickelte, theoretisch abgeleitete Kompetenzmodell der Modellkompetenz (Tab. 1) weist fünf Teilkompetenzen zur Beschreibung der Modellkompetenz auf, die jeweils in drei Niveaus ausdifferenziert werden, welche verschiedene Perspektiven auf (natur-)wissenschaftliche Modelle beschreiben.

**Tab. 1:** Teilkompetenzen und Niveaus des Kompetenzmodells der Modellkompetenz nach Upmeier zu Belzen & Krüger (a.a.O., vgl. auch Krell et al. 2016)

		KOMPLEXITÄT		
		Niveau I	Niveau II	Niveau III
TEILKOMPETENZ	Eigenschaften von Modellen	Modelle sind Kopien von etwas	Modelle sind idealisierte Repräsentationen von etwas	Modelle sind theoretische Rekonstruktionen von etwas
	Alternative Modelle	Unterschiede zwischen den Modellobjekten	Ausgangsobjekt ermöglicht Herstellung verschiedener Modelle von etwas	Modelle für verschiedene Hypothesen
	Zweck von Modellen	Modellobjekt zur Beschreibung von etwas einsetzen	Bekannte Zusammenhänge und Korrelationen von Variablen im Ausgangsobjekt erklären	Zusammenhänge von Variablen für zukünftige neue Erkenntnisse voraussagen
	Testen von Modellen	Modellobjekt überprüfen	Parallelisieren mit dem Ausgangsobjekt; Modell von etwas testen	Überprüfen von Hypothesen bei der Anwendung; Modell für etwas testen
	Ändern von Modellen	Mängel am Modellobjekt beheben	Modelle als Modell von etwas durch neue Erkenntnisse oder zusätzliche Perspektiven revidieren	Modell für etwas aufgrund falsifizierter Hypothesen revidieren

Anmerkung: Grau unterlegt sind die beiden Teilkompetenzen „Eigenschaften von Modellen“ und „Zweck von Modellen“, auf denen der vorliegende Beitrag den Fokus legt.

Niveau I und II bilden dabei in diesem Kompetenzmodell der Modellkompetenz eher eine mediale Perspektive auf Modelle ab, beschreiben also Reflexionen über das konkrete Modellobjekt (Niveau I) beziehungsweise über das Beziehungsgeflecht zwischen Modell und Original (Niveau II). Niveau III bezieht sich indes vermehrt auf die methodische Perspektive auf Modelle als

Instrumente im wissenschaftlichen Erkenntnisprozess, etwa zur hypothesenbasierten Vorhersage neuer Erkenntnisse (vgl. Upmeyer zu Belzen & Krüger a.a.O.).

### 3 Forschungsstand zur Modellkompetenz im Sachunterricht

Zum Sachunterricht und seiner Didaktik ist festzuhalten, dass in Bezug auf die Modellkompetenz von Grundschulkindern bislang nur wenige Forschungsansätze und dementsprechend -ergebnisse vorliegen. Vor dem Hintergrund der hohen unterrichtspraktischen Bedeutsamkeit von Modellen für das Lehren und Lernen im Sachunterricht ist eine Auseinandersetzung mit dem Thema essenziell. So werden beispielsweise in vielen Lehrwerken zum Sachunterricht Modelle zu Kreisläufen (z.B. Wasserkreislauf oder Recycling) abgebildet, und es wird damit implizit vorausgesetzt, dass die Lernenden die darin enthaltenen Informationen (z.B. Pfeile, mit denen der Verlauf des – gasförmigen – Wassers dargestellt wird) korrekt interpretieren. Hier ist bislang jedoch noch völlig unklar, inwieweit die Schüler/innen tatsächlich in der Lage sind, den fachlichen Gehalt dieser Modelle zu verstehen und/oder diese sinnvoll zu nutzen (z.B. zur Hypothesenbildung).

Qualitative Studien aus dem Sekundarbereich zeigen grundsätzlich, dass Schüler/innen in der Altersspanne von zehn bis 16 Jahren häufig wenig elaborierte Vorstellungen von Modellen haben (Grosslight et al. 1991, Grünkorn, Lotz & Terzer 2014). Für den Primarbereich deuten die wenigen vorliegenden Forschungsarbeiten auf ein überwiegend naives Modellverstehen von Grundschul/innen (vgl. Trier 2013) sowie auf eine Kontextabhängigkeit beim Verstehen von Aussagen über Modelle hin (vgl. Gogolin et al. 2017, Böschl et al. 2018).

Einige Studien aus dem anglo-amerikanischen Raum zeigen allerdings, dass Grundschulkindern durchaus in der Lage sind, Modelle zu konstruieren und in (sach-)unterrichtlichen Situationen anzuwenden und zu evaluieren. Voraussetzung dafür ist jedoch, dass der Unterricht Gelegenheiten zum Anwenden dieser Denk-, Arbeits- und Handlungsweise bereitstellt und die Lehrkraft – insbesondere leistungsschwächere Schüler/innen – durch gezielte Maßnahmen (z.B. durch einen Leitfaden für Modellevaluation) unterstützt (z.B. Vo et al. a.a.O., Forbes, Zangori & Schwarz 2015, Schwarz et al. 2009, Schwarz et al. 2012). Auch Studien aus der Entwicklungspsychologie mit dem Fokus auf naturwissenschaftliche Kompetenzen zeigen, dass Grundschulkindern nicht nur über konzeptuelles Fachwissen verfügen, sondern auch im Sinne des wissenschaftlichen Denkens beispielsweise ein Verständnis über die Logik des Testens (Sodian et al. 2002) und des Experimentierens (Toth, Klar &

Chen 2000) oder der Interpretation von Daten (Koerber et al. 2015) erwerben können bzw. bereits über dieses verfügen. Deutlich zeigen die Ergebnisse dieser Studien auch, dass die Ausprägungen dieser Fähigkeiten und Fertigkeiten nicht an ein bestimmtes Alter gebunden sind, sondern dass sich Entwicklungsveränderungen häufig in spezifischen Wissensdomänen und in Abhängigkeit von Förderangeboten vollziehen (Bullock, Sodian & Koerber 2009, Sodian 2004).

Die positiven Zusammenhänge zwischen dem Erwerb von Modellkompetenz und weiteren Zieldimensionen des Sachunterrichts, wie z.B. dem fachlichen Lernen, dem Aufbau eines elaborierteren Wissenschaftsverständnisses und von Interessen, sind bislang lediglich theoretisch postuliert (vgl. Lange & Hartinger a.a.O.) und nicht empirisch untersucht (vgl. Giere, Bickle & Mauldin 2006). Zudem ist festzuhalten, dass für die für den Sachunterricht relevante Altersgruppe weder etablierte theoretische Modelle der Modellkompetenz vorliegen, noch dass es valide Aussagen dazu gibt, wie die Modellkompetenz von Grundschüler/innen ausgeprägt ist. Es existieren jedoch theoretische Rahmungen und Erfassungsmethoden aus a) benachbarten Fachdidaktiken, die für eine andere Altersstufe entwickelt wurden (vgl. Krell et al. 2016) und b) der Entwicklungspsychologie für die grundschulrelevante Altersgruppe, die aber nicht spezifisch auf Modellkompetenz ausgerichtet sind (Koerber et al. 2011). Somit ergibt sich für die fachdidaktische Forschung die Aufgabe, diese Ansätze zu nutzen und auf die Anforderungen des Schulfachs Sachunterricht und seiner Didaktik zu übertragen.

#### **4 Konkretisierung der Fragestellung**

In unserer Untersuchung zur Modellkompetenz soll langfristig und übergeordnet folgender Frage nachgegangen werden: Kann Modellkompetenz im Sachunterricht der Grundschule gefördert werden? Daran anschließend gilt es zu untersuchen, ob sich – im Sinne der Kompetenzorientierung im Sachunterricht (GDSU 2013, Adamina 2013) – Zusammenhänge zwischen dem Lernen im Bereich der fokussierten Denk-, Arbeits- und Handlungsweisen und dem fachlichen Lernen von (perspektivenbezogenen) Inhaltsbereichen feststellen lassen. Damit sind übergeordnet dann auch Fragen nach der Gestaltung von Sachunterricht zur Förderung fachspezifischer Ziele tangiert, was ein genuin sachunterrichtsdidaktisches Forschungsfeld darstellt. Wenn man solch verallgemeinerbare Aussagen über die Wirkungen von Unterrichtsmaßnahmen oder über die Zusammenhänge einer allgemeinen Kompetenz zu perspektivenbezogenen Kompetenzziele erheben und beschreiben will, verlangt das ein empirisch-quantitativ ausgerichtetes Design, um hypothesentestend Ver-

gleiche mit Kontrollgruppen durchführen zu können. Dazu ist es allerdings wiederum erforderlich, Testinstrumente zu entwickeln, mithilfe derer man in der Lage ist, die Ausprägung der abhängigen Variable (in unserem Fall die Modellkompetenz von Grundschüler/innen) valide und zugleich ökonomisch zu erheben (Gogolin & Krüger 2016). Die Umsetzung solcher Forschungsdesigns zur Beantwortung der Forschungsfragen macht Instrumente erforderlich, die effizient auszuwertende Daten liefern, was vor allem durch Likert-Skalen, multiple bzw. selected-choice-Alternativen erreicht werden kann. Diese Verfahren erlauben zugleich auch ein hohes Maß an Objektivität. Erforderlich ist dann außerdem die Überprüfung der Reliabilität (Messgenauigkeit), die anhand der erhobenen Daten mit Hilfe von Koeffizienten wie Cronbach's Alpha berechnet wird und bei ausreichender theoretischer Klarheit und durch Maßnahmen der Pilotierung zumeist gesichert werden kann. Die Frage nach der Validität von Messwertinterpretationen wird indes als die wichtigste und kritischste Frage in Bezug auf die psychometrische Qualität eines Testinstruments gesehen (vgl. z.B. AERA, APA & NCME 2014). Im Gegensatz zu der traditionell verbreiteten Ansicht, Validität wäre eine Eigenschaft des Tests an sich (vgl. Rost 2004), wird sie in aktuellen Standards (z.B. Standards for Educational and Psychological Testing, vgl. AERA, APA & NCME 2014) als Eigenschaft der Testwerte (Kane 2013) definiert. Für jede Schlussfolgerung aus einem Testergebnis muss bezogen auf das Untersuchungsdesign (u.a. Testinstrumente) der Nachweis erbracht werden, dass sie tatsächlich gerechtfertigt ist. Von zentraler Bedeutung ist es deshalb, dass im Laufe der Testentwicklung und -erprobung ausreichend plausible, theoretische sowie empirische Evidenzen gesammelt werden, die eine valide Interpretation der Messwerte stützen (vgl. AERA et al. a.a.O., Gogolin 2017). Folglich sollte der Frage nach der Validität eine besondere Aufmerksamkeit gewidmet werden, weshalb dies auch einen Fokus des vorliegenden Beitrags darstellt.

Mit Blick auf unsere Forschungsarbeiten zur Modellkompetenz von Grundschüler/innen stellt sich die Frage, ob das von uns gewählte methodische Vorgehen, in dem theoretische Grundlagen aus benachbarten Fachdidaktiken übernommen und für Grundschulkindern adaptiert werden, zu validen Messwertinterpretationen führen kann. Um einen ersten Hinweis darauf zu sammeln, sind wir in unseren Forschungsarbeiten der Frage nachgegangen, ob Grundschüler/innen die Aussagen in den von uns auf der Basis von Arbeiten aus benachbarten Fachdidaktiken (aus dem MINT-Bereich) entwickelten Testinstrumenten wie theoretisch intendiert verstehen bzw. welche Verstehensprobleme sich in den Interpretationen der Aufgaben durch die Lernenden

zeigen. Bezogen auf unsere Studie und den vorliegenden Artikel lässt sich diese Frage folgendermaßen konkretisieren:

Sind die im Fragebogen erfassten Angaben der Kinder den verschiedenen Niveaus der Modellkompetenz valide zuzuordnen?

Für den vorliegenden Beitrag, in dessen Fokus die eingesetzten Forschungsmethoden stehen, soll zudem spezifisch folgender Frage nachgegangen werden: Ist das zur Beantwortung der Forschungsfrage gewählte Mixed-Methods Design geeignet, um Hinweise auf die Validität der Messwertinterpretationen zu erhalten?

## 5 Methodik der Forschungsarbeit

### 5.1 Methodologisches Prinzip

Für die Entwicklung unserer Testinstrumente greifen wir auf ein *Convergent-Mixed-Methods-Design* (Creswell & Plano Clark 2011) zurück. Der Grundgedanke dieses Designs sieht eine wissenschaftstheoretisch reflektierte Verschränkung vor, das heißt eine Kombination und Integration von qualitativen und quantitativen Methoden (Döring & Bortz 2016). In der von uns gewählten Variante des Mixed-Methods-Designs besitzen dabei beide Stränge (qualitativ und quantitativ) gleiche Priorität; es ist demnach kein Vorgehen dem anderen über- beziehungsweise untergeordnet (Kuckartz 2014). Ziel ist die Annäherung an die aufgeworfene(n) Forschungsfrage(n) durch die Zusammenführung (*convergence*) sämtlicher erhobener Daten durch die beiden miteinander verzahnten Stränge (Creswell & Plano Clark a.a.O.). Dies ermöglicht es, Abweichungen beziehungsweise Annäherungen der jeweiligen Datensätze und Ergebnisse vom beziehungsweise an das zu erfassende Konstrukt – hier die Modellkompetenz von Grundschüler/innen – umfassender sichtbar zu machen und zu beschreiben (Kuckartz a.a.O., Döring & Bortz a.a.O.). Ein vergleichbarer Ansatz wurde von Gogolin & Krüger (2016) für die Überprüfung von Aufgaben zum Modellverstehen in der Sekundarschule genutzt.

### 5.2 Erkenntnistheoretisch-methodologische Position und Begründung

In der vorliegenden Studie werden theoretische Annahmen und Erhebungsverfahren sowohl aus benachbarten Fachdidaktiken (aufgrund der Forschungslage v.a. aus der Didaktik der Biologie) sowie aus der pädagogischen Psychologie berücksichtigt. Dabei werden die theoretischen Modelle und Instrumente nicht ohne weiteres übernommen. Es ist dennoch anzunehmen, dass die national und international in benachbarten Fachdidaktiken für höhere Altersstufen entwickelten theoretischen Modelle der Modellkompetenz (vgl.

Upmeier zu Belzen & Krüger 2010) ein Fundament und damit den theoretischen Ausgangspunkt für Forschungsarbeiten hierzu im Grundschulbereich bilden können. Daher sollen quantitative Ansätze gewählt werden, um hypothesentestend den Stand der Forschung für die Grundschule weiter zu erarbeiten. Gleichzeitig ist davon auszugehen, dass diese Modelle und Rahmungen nicht unverändert und unüberprüft für den Sachunterricht der Grundschule übertragen werden können; vielmehr sind sie mit Blick auf die Kompetenzen und den Entwicklungsstand von Grundschulkindern zu adaptieren, auszudifferenzieren und zu schärfen. Es müssen also hypothesengenerierende Verfahren zum Einsatz kommen, um die theoretischen Modelle weiter zu entwickeln. Analog dazu wird die Entwicklung von Testinstrumenten zur Kompetenzerfassung auf Vorgehensweisen und Befunden der pädagogischen Psychologie aufgebaut, da man sich dort bereits seit Jahrzehnten mit der Erfassung von Kompetenzen beschäftigt (vgl. z.B. Sodian 2004, Sodian et al. 2002). Die Entwicklung von Testinstrumenten zur Erfassung von Modellkompetenz geschieht vor diesem erkenntnistheoretisch-methodologischen Hintergrund zu zweierlei Zwecken: Zum einen werden reliable und valide interpretierbare Testinstrumente benötigt, um in empirisch-quantitativ ausgerichteten Studien unter dem Prozess-Produkt-Paradigma (vgl. Köller 2008) Kompetenzen zu beschreiben und zu erfassen beziehungsweise um Wirkungen von Interventionen zu testen. Zum anderen sollen über den Einsatz der Testinstrumente Hinweise auf die Ausprägung der Modellkompetenz bei Grundschüler/innen gesammelt und die theoretischen Modelle dadurch auch für den Bereich der Grundschule ausgeschärft werden.

### **5.3 Methoden der Datengewinnung und Datenerhebung**

Es gibt bewährte Verfahren aus der Sekundarstufe mit theoretischem Bezug auf das zuvor beschriebene Kompetenzmodell, mit deren Hilfe Modellverstehen erhoben wurde (vgl. z.B. Gogolin & Krüger 2016, 2017; Krell et al. 2016, Grünkorn et al. 2014). Durch die Adaption dieser Verfahren auf die Grundschule (Gogolin et al. 2017) und vor allem durch die Überprüfung der Validität ergeben sich zugleich Hinweise auf die Fähigkeit von Grundschüler/innen, Modelle zu verstehen. Weitere Befunde (vgl. u.a. Gogolin & Krüger 2018; Krell, Upmeier zu Belzen & Krüger 2014, Sodian et al. 2002) legen zudem nahe, dass Kontexte Einfluss auf die Performanz von Lernenden haben. Dieser Befund ist sowohl für die Einschätzung der Validität aber auch mit Blick auf die Verwendung von Modellen in den verschiedenen Perspektiven des Sachunterrichts von großer Bedeutung. Für den Einsatz im Sachunterricht der Grundschule bieten sich zunächst komplexitätsreduzierte und

wissensarme Aufgaben sowie eine Einbettung der Aufgabenkontexte in Geschichten an (vgl. Akerson & Donnelly 2010, Sodian et al. 2002).

Um zu erheben, ob die Fragen des Fragebogens von den Schüler/innen wie intendiert verstanden wurden und die Angaben der Kinder von daher den intendierten Niveaus zuzuordnen sind, wurden in enger Anlehnung an das theoretisch abgeleitete Kompetenzmodell der Modellkompetenz (Upmeier zu Belzen & Krüger 2010) zentrale Komponenten bezüglich der dort postulierten Teilkompetenzen erhoben. Wir berichten im Folgenden über zwei Teilkompetenzen, die für das Arbeiten im Sachunterricht von besonderer Bedeutung sind: „Eigenschaften von Modellen“ und „Zweck von Modellen“ (vgl. auch Gogolin et al. 2017, Böschl et al. 2018).

Zur Erhebung des Verstehens der jeweils für diese beiden Elemente formulierten Niveaus wurden in einem quantitativen Ansatz Aufgaben zu fünf unterschiedlichen Kontexten entwickelt. Diese Kontexte sind so gewählt, dass neben dem naturwissenschaftlichen Bereich (Wasserkreislauf) auch geographische (Erdinneres), erdzeitgeschichtliche (Dinosaurier), biologische (Fischatmung) sowie technische (Fahrradantrieb) Inhalte aufgegriffen werden. Die zu jedem Kontext konstruierten Antwortalternativen wurden in zufälliger Reihenfolge in Ratingaufgaben zusammengestellt. Jede dieser Aufgaben besteht aus einem standardisierten kontextspezifischen Aufgabestamm, der für die Lernenden als kurze Geschichte formuliert ist, und einer Abbildung. Über die fünf Kontexte und die zwei hier fokussierten Teilkompetenzen hinweg handeln diese Geschichten abwechselnd simplifizierend entweder von einer Gruppe von Forscherinnen oder einer Gruppe von Forschern (s. Abb. 1):

### Der Brachiosaurus

Viele Forscher beschäftigen sich mit Brachiosauriern. Das waren sehr große Dinosaurier, die vor langer Zeit lebten. Sie haben sich wahrscheinlich von Pflanzen ernährt.

Vor kurzem war eine Gruppe von Forschern zu Besuch in einem Museum. Dort haben sie den echten Schädel eines Brachiosaurus gesehen und ein Foto gemacht. Dieses Foto kannst du hier sehen:

Die Forscher interessieren sich dafür, wie Brachiosaurier gekaut haben. Ein anderer Wissenschaftler hat ihnen den Tipp gegeben, ein Modell herzustellen. Die Forscher sind sich aber nicht einig,

wofür das Modell dann genutzt werden soll.



**Abb. 1:** Auszug aus Aufgabe zur Erhebung der Teilkompetenz „Zweck von Modellen“ im Kontext „Dinosaurier“ (Teil 1): Abbildung des Originals (Brachiosaurusschädel) und kontextspezifischer Aufgabenstamm, eingebettet in eine kurze Geschichte.

Anschließend folgen ein standardisierter Impuls für die Aufgabenbearbeitung und jeweils drei Antwortalternativen. Letztere entsprechen jeweils einem Niveau des Kompetenzmodells der Modellkompetenz nach Upmeyer zu Belzen und Krüger (2010). Zum Ende der Aufgabe wurden die Schüler/innen in einem Forced-Choice Format (vgl. Gogolin & Krüger 2016, Moosbrugger & Kelava 2012) um die Auswahl der ihrer Ansicht nach besten Aussage gebeten (s. Abb. 2). (Um einen vereinfachten Zugang zum Beispiel in Abbildung 2 zu ermöglichen, sind die nachfolgend dargestellten Antwortalternativen – anders als in den tatsächlichen Testheften – nicht in zufälliger Reihenfolge abgebildet, sondern hierarchisch angeordnet – die Aussage des ersten Forschers entspricht dem Niveau I der dort dargestellten Teilkompetenz „Zweck von Modellen“, die Aussage des zweiten Forschers dem Niveau II und die Aussage des dritten Forschers dem Niveau III.)

**Herr Müller sagt:**

„Modelle sollen zeigen, wie etwas aussieht oder funktioniert. Mit einem Modell des Brachiosaurus können wir anderen zeigen, welche Zähne Brachiosaurier beim Kauen nutzen.“

**Wie sehr stimmst du der Meinung von Herrn Müller zu? Kreuze an!**

gar nicht	kaum	mittelmäßig	ziemlich	völlig
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Herr Schneider sagt:**

„Modelle sollen erklären, wie etwas zusammenhängt. Mit einem Modell des Brachiosaurus können wir anderen erklären, welchen Einfluss bestimmte Zähne beim Kauen hatten.“

**Wie sehr stimmst du der Meinung von Herrn Schneider zu? Kreuze an!**

gar nicht	kaum	mittelmäßig	ziemlich	völlig
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Herr Schmidt sagt:**

„Modelle sind dazu da, um Vermutungen aufzustellen. Mit einem Modell des Brachiosaurus können wir eigene Vermutungen darüber aufstellen, wie Brachiosaurier wahrscheinlich gekaut haben.“

**Wie sehr stimmst du der Meinung von Herrn Schmidt zu? Kreuze an!**

gar nicht	kaum	mittelmäßig	ziemlich	völlig
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Wem stimmst du am meisten zu? Mache ein Kreuz!**

Herr Müller	Herr Schneider	Herr Schmidt
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Abb. 2:** Auszug aus Aufgabe zur Erhebung der Teilkompetenz „Zweck von Modellen“ im Kontext „Dinosaurier“ (Teil 2): standardisierter Impuls für Aufgabebearbeitung und drei kontextspezifische Antwortalternativen auf unterschiedlichen Teilkompetenzniveaus (I-III) und abschließende Forced-Choice Aufgabe.

Insgesamt bewerteten  $N=110$  Grundschüler/innen der Klassenstufen drei und vier auf einer fünfstufigen, verbaläquidistanten Likert-Skala (gar nicht – kaum – mittelmäßig – ziemlich – völlig; vgl. Moosbrugger & Kelava a.a.O.), wie sehr die einzelnen Antwortalternativen pro Aufgabe ihrer eigenen Ansicht entsprechen. Keiner der untersuchten Lernenden hatte zuvor bereits expliziten Unterricht über Modelle. Für den qualitativen Ansatz wurden im Anschluss an die schriftliche Befragung mit 63 der 110 Schüler/innen voll-

strukturierte Einzelinterviews geführt. Hierfür unterstützte der jeweils eigene, zuvor ausgefüllte Fragebogen als *Stimulated Recall* die Erinnerung an das Bearbeiten der Aufgaben (Sandmann 2014). Im Interview las die interviewende Person der jeweiligen Schülerin bzw. dem jeweiligen Schüler zunächst zur ersten Aufgabe die erste Antwortalternative vor. Dabei betrachteten Kind und Interviewer(in) die Bewertung des Kindes (auf der Likert-Skala und dem Forced-Choice Teil). Anschließend wurde es gebeten, seine Bewertung zu begründen. Das Vorgehen wurde für alle drei Antwortalternativen durchgeführt.

#### **5.4 Datenanalyse**

Zur Auswertung der Daten aus dem quantitativen Strang wurden die fünfstufigen, verbaläquidistanten Likert-Skalen sowie die im Forced-Choice Teil gesetzten Kreuze tabellarisch erfasst und kodiert (Likert-Skala: 1 = gar nicht; 2 = kaum; 3 = mittelmäßig; 4 = ziemlich; 5 = völlig; Forced-Choice Teil: 1 = Aussage Niveau I; 2 = Aussage Niveau II; 3 = Aussage Niveau III).

Die Auswertung der Daten aus dem qualitativen Strang erfolgte im Anschluss an die Transkription der Audiodateien mithilfe der evaluativen Inhaltsanalyse (Mayring 2010). Erhoben wurde, inwieweit die Schülerin bzw. der Schüler die Antwortalternativen so verstanden hatte, wie sie intendiert waren bzw. welche Verstehensprobleme sich in den Interpretationen der Aufgaben durch die Lernenden zeigten (vertiefend hierzu Gogolin et al. 2017). Die Interviewtranskripte wurden mit dem Programm MAXQDA mittels eines entwickelten Kodierleitfadens in drei Kategorien kodiert: 1 = intendiertes Niveau verstanden; 0 = intendiertes Niveau nicht verstanden; 3 = Antwort nicht aussagekräftig (s. Tab. 2).

**Tab. 2:** Kategorisierte Beispielaussagen zur Teilkompetenz „Zweck von Modellen“ (Niveau I, II, III) im Kontext „Dinosaurier“

	KATEGORIE	BEISPIELAUSSAGEN
Intendiertes Niveau verstanden	N I	Also er [Herr Müller] sagt, dass man mit einem Modell, wenn man z.B. einen Kopf so gesagt als Modell hat, dass man dann, eben auch anderen Leuten, die sich dafür interessieren, auch zeigen kann, mit welchen Zähnen der Brachiosaurus kaut.
	N II	Also ich hatte da auch ziemlich angekreuzt, weil ja es soll ja eigentlich wie gesagt auch erklären, wie der Brachiosaurus oder auch Brontosaurus jetzt gekaut hat, und bei so einem Modell wie gesagt kann man ja eigentlich alles so sehen, also man so etwas jetzt gut herstellt, dann könnte man ja eigentlich sehr gut sehen, und da hab ich dann auch ziemlich angekreuzt (...).
	N III	Mit dem Modell Vermutungen aufstellen kann, kann man sich ja überlegen, wie es am sinnvollsten ist (...) und dann ausprobieren, ob das so geht, alles.
Intendiertes Niveau nicht verstanden	N I	Dass/sie wollen eben einen Kopf, wollen sie so machen, dass sie es anderen erklären können, also dass sie auch bewegen können. Aber dann ist es ja kein richtiger Kopf, weil die dann ja auch noch andere Knochen oder so reingebaut haben. Und deswegen kauen, denk ich immer.
	N II	Also das versteh ich auch sehr gut, und ich find es auch sinnvoll, weil, andere Leute wollen ja auch sehen, was/wie so ein Dinosaurier früher gegessen und gekaut hat.
	N III	Dort habe ich mittelmäßig angekreuzt, weil also ich finde irgendwie, das sind alles dieselben Antworten sind und ich konnte mich da nicht immer so entscheiden. Aber Vermutungen aufzustellen, das war jetzt so etwas mittelmäßig von mir, weil die anderen Antworten fand ich so etwas besser. Kann man sagen.
Antwort nicht aussagekräftig	N I	Also ich hab bei dem ziemlich angekreuzt, weil das war auch schon so etwas besser, weil ich find so Modelle eigentlich auch sehr cool. Und dort hätte ich dann ziemlich angekreuzt, weil das schon so mehr zum ja mehr passte, kann man so zu sagen sagen.
	N II	Das (...) die Forscher Menschen erklären wollen, die aus anderen Ländern kommen, vielleicht und (...) dann können die das (unv.) Herrn Schneider sagen, (...) welchen Einfluss die beim Kauen hatten.
	N III	Also ich finde Herr, als das was Herr Schmidt gesagt hat (...), mittelmäßig, weil, z.B., weil (...) weil man weiß ja nicht (...), wie die wirklich, also wirklich gekaut haben (...) und ja, weiter fällt mir nichts ein.

Um die Interrater-Reliabilität zu überprüfen, wurde eine Teilstichprobe (10 von  $N=110$ ) zweikodiert. Mit einem berechneten Cohens Kappa von  $\kappa \geq 0,74$  (vgl. Wirtz & Caspar 2002) ist diese als gut einzuschätzen.

Weiterführend ermöglichte das gewählte Convergent-Mixed-Methods-Design (Creswell & Plano Clark 2011) durch die Zusammenführung der erhobenen qualitativen und quantitativen Daten die Analyse eines Sub-Samples. Dieses wurde aus den kodierten Aussagen der Schüler/innen (qualitative Daten) und den ausgefüllten Fragebögen mit Likert-Skalen und Forced-Choice Aufgaben (quantitative Daten) extrahiert. In das Sub-Sample gehen die Aussagen ein,

die in drei ausgewählten Kontexten (Erdinneres, Fischatmung, Fahrradantrieb) in Abhängigkeit vom jeweiligen Niveau und der anvisierten Teilkompetenz als „wie intendiert verstanden“ kodiert wurden. Die Analyse soll neue Einsichten bezüglich der Ausprägung eines Niveaus III im Primarbereich erbringen.

Für den Vergleich der Unterschiede in den Verteilungen bezüglich der Kompetenzniveaus bzw. Kontexte wurden aufgrund des kategorialen Niveaus der Daten Chi-Quadrat Tests durchgeführt.

## 6 Ergebnisse

Durch den Bezug der Interviewdaten zu den Angaben des Fragebogens zeigte sich, dass die Antwortalternativen bei den ausgewählten Kontexten auf den Niveaus I und II von mehr als der Hälfte der Schüler/innen wie theoretisch intendiert verstanden wurden ( $M_{NI} = 60\%$ ,  $M_{NII} = 51\%$ ,  $M_{NIII} = 26\%$ ;  $\chi^2_{\text{(Eigenschaften)}} = 2,43$ ,  $p = 0,29^*$ ;  $\chi^2_{\text{(Zweck)}} = 6,77$ ,  $p = 0,03^{**}$ ) (zu weiteren Details vgl. Böschl et al. 2018). Die Ausnahme bilden dabei die Antwortalternativen der Kontexte „Wasserkreislauf“ und „Fischatmung“ in der untersuchten Teilkompetenz „Zweck von Modelle“. Die Aussagen, die zu diesen Kontexten von uns formuliert worden waren, um das jeweilige Niveau angemessen zu versprachlichen, wurden insbesondere auf Niveau II vom Großteil der Lernenden nicht so verstanden, wie von uns intendiert, bzw. die Antworten waren nicht aussagekräftig. Dennoch ist über alle Kontexte hinweg der Trend zu beobachten, dass das Verstehen mit zunehmendem Niveau, insbesondere aber hinsichtlich des Niveaus III, abnimmt (vgl. a.a.O.).

Betrachtet man die Aussagen der Schüler/innen auf rein quantitativer Basis so zeigt sich (vgl. Tab. 3), dass die Zustimmung der Schüler/innen zu den Niveau I-Aussagen auf den Likert-Skalen über beide Teilkompetenzen (mit einer Ausnahme) hinweg höher ausfiel als auf den beiden anderen Niveaus. Zudem wählten die Lernenden auch in den Forced-Choice-Items häufiger Niveau I als die anderen Niveaus.

**Tab. 3:** (links) Mittelwerte der in den Aufgaben gekreuzten Likert-Skalen Items für die Aussagen auf den Niveaus I, II und III (1= gar nicht; 2= kaum; 3= mittelmäßig; 4= ziemlich; 5= völlig); (rechts) Angabe der Häufigkeit der gewählten Aussagen in den Forced-Choice Items auf den Niveaus I, II und III.

KONTEXT	TEILKOMPETENZ: EIGENSCHAFTEN VON MODELLEN					
	Likert-Skalen Mittelwerte			Forced-Choice-Item Häufigkeit		
	N I	N II	N III	N I	N II	N III
Dinosaurier	3,75	2,89	3,06	17	8	10
Wasserkreislauf	4,16	2,68	3,05	10	3	5
Fahrradantrieb	3,43	2,88	3,29	3	1	3
Fischatmung	4,20	2,25	2,56	5	3	1
Erdinneres	3,70	3,10	3,10	4	2	4
KONTEXT	TEILKOMPETENZ: ZWECK VON MODELLEN					
	Likert-Skalen Mittelwerte			Forced-Choice-Item Häufigkeit		
	N I	N II	N III	N I	N II	N III
Dinosaurier	3,92	3,79	3,69	21	15	18
Wasserkreislauf	4,32	3,68	3,83	16	4	13
Fahrradantrieb	4,14	4,00	3,07	9	5	0
Fischatmung	3,57	4,00	3,86	2	2	3
Erdinneres	3,60	3,50	3,40	5	2	3

Bemerkenswert ist zudem, dass die Zustimmung zu den Niveau III Aussagen – insbesondere zur Teilkompetenz „Eigenschaften von Modellen“ – durchschnittlich höher ist als für die Aussagen auf Niveau II (s. Tab. 3), obwohl das Niveau III von den wenigsten Schüler/innen wie intendiert verstanden wurde (s. Abb. 3; vgl. Böschl et al. 2018).

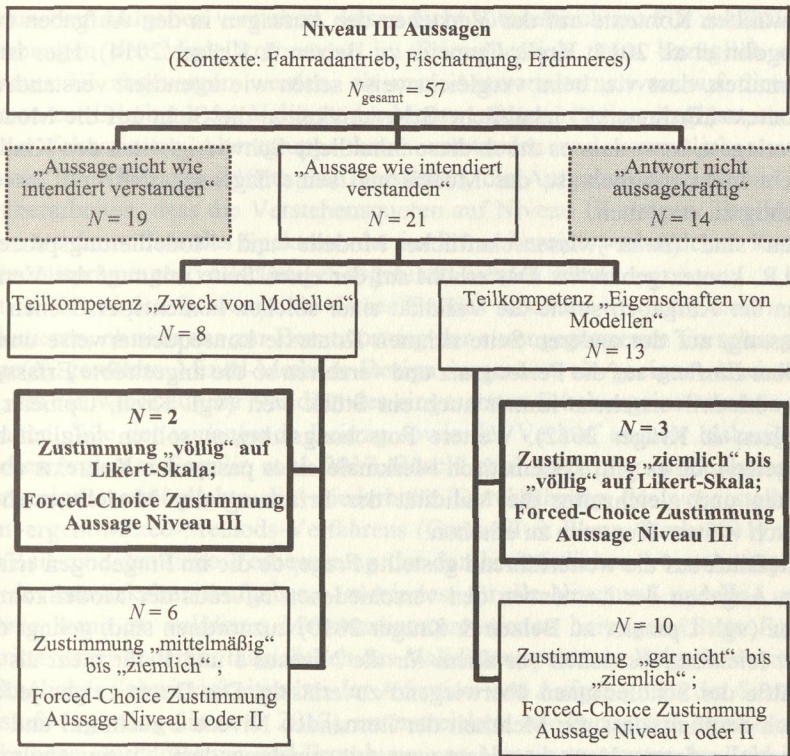
Betrachtet man dem gegenüberstellend das aus der Zusammenführung der quantitativen und qualitativen Daten – zunächst für drei ausgewählte Kontexte – synthetisierte Sub-Sample an Aussagen ( $N=98$ ), die „wie intendiert verstanden“ wurden, so zeigt sich (vgl. Tab. 4), dass auch hier die durchschnittliche Zustimmung der Lernenden zu den Aussagen auf Niveau I auf den Likert-Skalen über beide Teilkompetenzen hinweg (mit zwei Ausnahmen) am höchsten ausfiel.

**Tab. 4:** Mittelwerte der in den Aufgaben gekreuzten Likert-Skalen Items für das Sub-Sample an „wie intendiert verstanden“ Aussagen ( $N=98$ ) auf den Niveaus I, II und III (1= gar nicht; 2= kaum; 3= mittelmäßig; 4= ziemlich; 5= völlig);

KONTEXT	TEILKOMPETENZ					
	Eigenschaften von Modellen			Zweck von Modellen		
	Likert-Skalen Mittelwerte			Likert-Skalen Mittelwerte		
	N I	N I	N I	N I	N II	N III
Fahrradantrieb	3,50	4,36	4,36	4,36	4,00	3,50
Fischatmung	4,38	4,17	4,17	4,17	4,00	0,00
Erdinneres	4,33	4,00	4,00	4,00	4,50	4,00

Auffällig ist bei der Betrachtung des Sub-Samples jedoch der Trend, dass nun – mit Ausnahme der Aussagen zum Kontext „Erdinneres“ – die durchschnittliche Zustimmung zu den Aussagen, anders als noch zuvor (s. Tab. 3), über die Kontexte und Teilkompetenzen mit zunehmendem Niveau abnimmt. Abgesehen von einer größeren Abweichung („Erdinneres“, Niveau I), liegen die vorliegenden durchschnittlichen Mittelwerte der Zustimmung zu den Aussagen auf den Likert-Skalen der Tabellen 3 und 4, gerade bei der untersuchten Teilkompetenz „Eigenschaften von Modellen“, relativ nah beieinander. Die Werte der Teilkompetenz „Zweck von Modellen“ weichen im direkten Vergleich hingegen mitunter deutlich voneinander ab, insbesondere bei den Kontexten „Fischatmung“ (Niveau III) und „Erdinneres“ (Niveau II; Niveau III).

Auch um einschätzen zu können, ob es für den Sachunterricht der Grundschule ein angemessenes Ziel sein kann, Kinder auf das Niveau III der Modellkompetenz zu führen, sollen hier noch die Ergebnisse der im Sub-Sample enthaltenen Niveau III-Aussagen analysiert werden (s. Abb. 3). So finden sich im relevanten Niveau III Sub-Sample ( $N=21$ ) über die erhobenen Teilkompetenzen hinweg fünf Untersuchungsteilnehmer/innen, deren Aussagen zugleich in der evaluativen qualitativen Inhaltsanalyse (Mayring 2010) in die Kategorie „Aussage wie intendiert verstanden“ kodiert worden waren, die den Items auf Niveau III in den Aufgaben den Likert-Skalen „ziemlich“ bis „völlig“ zustimmten und bei den Forced-Choice-Items Niveau III am meisten zustimmten (jeweils quantitative Auswertung).



**Abb. 3:** Ergebnisse der vertiefenden Analyse der Niveau III Aussagen ( $N=57$ ) mit Fokus auf die Kategorie „Aussage wie intendiert verstanden“ ( $N=21$ )

## 7 Diskussion und Ausblick

Die vorliegende Studie diente unter anderem dazu, Hinweise zur Frage zu sammeln, ob Kinder im Grundschulalter die Aussagen in den Messinstrumenten zur Erfassung von Modellkompetenz, die von uns auf Basis der Arbeiten aus benachbarten Fachdidaktiken entwickelt bzw. adaptiert wurden, wie theoretisch intendiert verstehen. Wie die Ergebnisse zeigen, erweist sich für den Grundschulbereich die theoretische Ableitung von Antwortalternativen aus dem skizzierten Kompetenzmodell der Modellkompetenz (vgl. Upmeyer zu Belzen & Krüger 2010, Krell et al. 2016) für die Niveaus I und II als durchaus tragfähige Basis für die Generierung von validen Aufgaben (vgl. Bullock et al. 2009, Koerber et al. 2016). Sehr deutlich ist der Einfluss der

gewählten Kontexte auf das Verstehen der Aussagen in den Aufgaben (vgl. Gogolin et al. 2017; Krell, Upmeyer zu Belzen & Krüger 2014). Hier ist zu vermuten, dass v.a. beim vergleichsweise selten wie intendiert verstandenen Kontext „Erdinneres“ inhaltliche Schwierigkeiten die Sicht auf die Modelle überlagert, bzw. dass es durch diese inhaltliche Schwierigkeiten den Kindern nicht mehr möglich ist, das Modell und seine Eigenschaften und Zwecke richtig zu verstehen.

Nun sind (natur-)wissenschaftliche Modelle und Modellierungsprozesse i.d.R. kontextgebunden. Das erhöht auf der einen Seite mitunter das Verstehen der Aufgaben sowie die Validität einer solchen kontextspezifischen Erfassung, auf der anderen Seite nehmen Kontexte konsequenterweise unmittelbar Einfluss auf die Performanz und verzerren so die angestrebte Erfassung des Modellverstehens immer auch ein Stück weit (vgl. Krell, Upmeyer zu Belzen & Krüger 2012). Weitere Forschungsarbeiten sollten folglich hier ansetzen, da es gilt, systematisch Merkmale eines passenden Kontexts abzuleiten und gleichzeitig die Validität der Erfassung des Modellverstehens durch weitere Kontexte zu erhöhen.

Mit Blick auf die weiterführend gestellte Frage, ob die im Fragebogen erfassten Angaben der Lernenden den verschiedenen Niveaus der Modellkompetenz (vgl. Upmeyer zu Belzen & Krüger 2010) zuzuordnen sind, gelingt dies mit leichten Abstrichen vor allem für die Niveaus I und II für mehr als die Hälfte der Schüler/innen überwiegend zuverlässig. Die Daten zeigen jedoch auch deutlich, dass die Mehrheit der Lernenden Niveau I zustimmt und das theoriekonforme Verstehen insgesamt mit zunehmendem Niveau abnimmt. Inhaltlich decken sich diese Ergebnisse damit mit den Erkenntnissen aus qualitativen Studien zur Modellkompetenz aus der Sekundarstufe, welche den Schüler/innen ein überwiegend naives Modellverstehen attestieren (z.B. Grünkorn et al. 2014, Grosslight et al. 1991).

Obwohl die mangelnde Passung zwischen theoretischer Intention und Interpretation bei Antwortalternativen auf Niveau III aus testtheoretischer Sicht weiterhin problematisch bleibt, ermöglicht die Zusammenführung der Ergebnisse der quantitativen und qualitativen Stränge dennoch inhaltlich interessante Einblicke bezüglich der Niveau III Ausprägung der Modellkompetenz von Grundschüler/innen. So finden sich in den analysierten Sub-Sample Aussagen (N=98) Indizien dafür, dass auch Grundschulkinder durchaus in der Lage sein können – analog zu Erkenntnissen der Entwicklungspsychologie (vgl. Sodian 2004, Sodian et al. 2002) – über höhere Niveaus als Niveau I in den untersuchten Ausprägungen (Zweck und Eigenschaften von Modellen) von Modellkompetenz nachzudenken und diese zum Teil auch vertreten zu können (s. Abb. 3). Dabei stimmen Lernende, die etwa die Niveau III Ausprägung

gen unserer Aufgaben verstehen, diesen nicht automatisch zu. Für zukünftige Erhebungen der Modellkompetenz im Primarbereich muss bezüglich des Niveaus III abgewogen werden, ob man Antwortalternativen auf diesem Niveau zur Erhöhung der Validität der Messwertinterpretation weglässt, da es in der fokussierten Altersgruppe kaum verstanden und vermutlich wenig präsent ist. Alternativ könnte man versuchen, die Aufgaben noch einmal so zu überarbeiten, dass die Verstehensquoten auf Niveau III steigen. Der Vorteil wäre dann, dass das Spektrum der im Primarbereich vorhandenen Kompetenzausprägungen vollständiger abgebildet werden würde.

Mit Blick auf den Schwerpunkt dieses Bandes möchten wir noch die Frage diskutieren, ob sich das zur Beantwortung der vorangegangenen Forschungsfrage(n) gewählte Mixed-Methods-Design insgesamt als geeignet erwiesen hat, um Hinweise auf die Validität des eingesetzten Erhebungsinstruments zu erhalten. Wie der vorliegende Beitrag sowie der Verlauf unserer bisherigen Studien (vgl. u.a. Gogolin et al. 2017, Böschl et al. 2018) zur Modellkompetenz zeigen, ermöglichte die wiederholte Durchführung des vorgestellten Convergent-Mixed-Methods-Verfahrens (Creswell & Plano Clark 2011) mit der Zeit eine deutliche Verbesserung der durchschnittlichen Verstehensquoten der entwickelten Aufgaben – zumindest für die Niveaus I und II – und konnte somit zu valideren Messwertinterpretationen beitragen. Stete Fortschritte, unter anderem hinsichtlich des Validitätsaspekts, in Folge bzw. trotz anfänglicher Schwierigkeiten bei der Adaption von Aufgaben aus der Sekundarstufe (vgl. Gogolin et al. 2017), unterstreichen, wie bedeutsam eine solche mehrfache Durchführung und Erprobung von Messinstrumenten im Feld ist. Mithilfe von Reliabilitäts- oder Faktorenanalysen wäre es nicht möglich gewesen, die mangelnde Passung zwischen theoretischer Intention und Interpretation der Lernenden bei Antwortalternativen insbesondere auf dem Niveau III aufzuzeigen (vgl. a.a.O.), da das Antwortverhalten über die verschiedenen Items – unabhängig vom Verständnis der Frage – durchaus reliabel war. Das gewählte Mixed-Methods-Design ist demnach ein zentraler „Baustein“, um bei der Untersuchung von Messinstrumenten und der komplexen Aufgabe sicher zu stellen, dass diese valide interpretiert werden können.

Durch das eingesetzte Verfahren konnten wir aber auch erkennen, dass weiterführende, rein quantitative Berechnungen und Interpretationen (noch) nicht sinnvoll möglich sind – insbesondere mit Blick auf das Niveau III. Grund hierfür ist, dass es (noch) weitgehend unklar ist, ob die Lernenden – mit wenigen möglichen Ausnahmen (vgl. Abb. 3) – die Aussagen beispielsweise nicht verstehen, weil sie nicht über eine entsprechende Ausprägung von Modellkompetenz auf diesem Niveau verfügen oder weil sie die Aussagen

sprachlich nicht verstehen. Das unerwartete Muster von mehr Zustimmung zu den Aussagen auf Niveau III als auf Niveau II – insbesondere bezüglich der Teilkompetenz „Eigenschaften von Modellen“ – in den durchschnittlichen, rein quantitativen Zustimmungsraten auf den Likert-Skalen (vgl. Tab. 3), kann indes ebenfalls durch die methodisch konsequente Zusammenführung der quantitativen und qualitativen Daten (vgl. Tab. 4, Abb. 3) erklärt werden. So stimmen zwar bei zunächst rein deskriptiver Betrachtung der quantitativen Daten mehr Grundschüler/innen den Aussagen auf Niveau III zu als jenen auf Niveau II, betrachtet man jedoch etwa das Sub-Sample an „wie intendiert verstandenen“ Aussagen, zeigt sich dieses Muster nicht mehr (vgl. Tab. 4). Dies kann mitunter darauf zurückzuführen sein, dass die Schüler/innen die Aussagen, die Niveau III repräsentieren sollten, anders verstanden haben. Ausgehend davon erscheinen auf den Niveaus I und II quantitative Analysen und entsprechende Interpretationen aber durchaus zulässig.

Zusammenfassend lassen sich mit den generierten und wiederholt durch das vorgestellte Convergent-Mixed-Methods-Verfahren überprüften Erhebungsmethoden erste inhaltliche Fragen – zunächst nach der Ausprägung des Modellverstehens von Grundschüler/innen bearbeiten. Somit wird ein Beitrag zur Theoriebildung innerhalb der Sachunterrichtsdidaktik geleistet, wobei zugleich die Anschlussfähigkeit an benachbarte Fachdidaktiken (MINT-Fächer der Sekundarstufe) sowie die Entwicklungspsychologie national und international gegeben ist. Um dabei weiterführend der für den Sachunterricht einleitend geforderten Kompetenzorientierung (GDSU 2013) gerecht zu werden, sehen wir perspektivisch die Notwendigkeit, neben dem Modellverstehen auch die Modellierungstätigkeiten in die Erfassung miteinzubeziehen. Hierfür sollen kompetenzorientierte Aufgaben (Konglomerat aus Wissen und Handlung; vgl. Adamina 2013) entwickelt werden. Orientiert am Vorgehen der vorgestellten Untersuchung (Theoriebezug; stete Rückkoppelung im Feld) soll dies in Erweiterung des Kompetenzmodells der Modellkompetenz von Upmeyer zu Belzen und Krüger (2010) letztlich zu präziseren theoretischen Rahmungen hinsichtlich der Modellkompetenz (vgl. Forbes, Lange-Schubert, Böschl & Vo im Druck) führen und damit weiter zur fortschreitenden Theoriebildung für den Sachunterricht der Grundschule und seine Didaktik beitragen.

## Literatur

- Adamina, M. (2013): Kompetenzorientierung im Sachunterricht am Beispiel der geographischen Perspektive. In: E. Gläser & G. Schönknecht (Hrsg.): Sachunterricht in der Grundschule: entwickeln – gestalten – reflektieren. Frankfurt: Grundschulverband, 171-187.

- Akerson, V. & Donnelly, L.A. (2010): Teaching Nature of Science to K-2 Students: What Understanding Can They Claim? In: *International Journal of Science Education*, 32, 1, 97-124.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (AERA, APA & NCME) (Eds.) (2014): *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Baur, A. (2015): Inwieweit eignen sich bisherige Diagnoseverfahren des Bereichs Experimentieren für die Schulpraxis? In: *Biologie Lehren und Lernen – Zeitschrift für Didaktik der Biologie*, 19, 1, 26-37. URL: <http://zdb.uni-bielefeld.de/index.php/zdb/article/view/316/302> [11.12.2017].
- Böschl, F.; Gogolin, S.; Lange-Schubert, K. & Hartinger, A. (2018): Modellverstehen von Grundschüler/innen in Abhängigkeit von Kontext und Kompetenzniveau. In: U. Franz; H. Giest; A. Hartinger; A. Heinrich-Dönges & B. Reinthoffer (Hrsg.): *Handeln im Sachunterricht – konzeptionelle Begründungen und empirische Befunde*. Bad Heilbrunn: Verlag Julius Klinkhardt, 93-100.
- Bullock, M.; Sodian, B. & Koerber, S. (2009): Doing experiments and understanding science: Development of scientific reasoning from childhood to adulthood. In: W. Schneider & M. Bullock (Eds.): *Human development from early childhood to early adulthood. Findings from the Munich Longitudinal Study*. New York, NY: Taylor & Francis Group LLC, 173-197.
- Clement, J. (2000): Model based learning as a key research area for science education. In: *International Journal of Science Education*, 22, 9, 1041-1053.
- Cresswell, J.W. & Plano Clark, V.L. (2011): *Designing and Conducting Mixed Methods Research*. Los Angeles, CA: SAGE.
- Döring, N. & Bortz, J. (2016): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer.
- Fleischer, J.; Koeppen, K.; Kenk, M.; Klieme, E. & Leutner, D. (2013): Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. In: *Zeitschrift für Erziehungswissenschaft*, 16, 1, 5-22.
- Forbes, C.T.; Zangori, L. & Schwarz, C.V. (2015): Empirical validation of integrated learning performances for hydrologic phenomena. 3rd-grade students' model-driven explanation-construction. In: *Journal of Research in Science Teaching*, 52, 7, 895-921.
- Forbes, C.T.; Lange-Schubert, K.; Böschl, F. & Vo, T. (im Druck): Supporting Primary Students' Developing Model Competence for Water Systems. In: A. Upmeyer zu Belzen, D. Krüger & J. van Driel (Eds.): *Towards a Competence-based View on Models and Modeling in Science Education. (Models and Modeling in Science Education)*. Cham: Springer International Publishing.
- Gesellschaft für Didaktik des Sachunterrichts (GDSU) (2013): *Perspektivrahmen Sachunterricht. Vollständig überarbeitete und erweiterte Auflage*. Bad Heilbrunn: Klinkhardt.
- Giere, R.; Bickle, J. & Mauldin, R. (2006): *Understanding scientific reasoning*. London: Thomson.
- Gilbert, J.K. & Justi, R. (2016): *Modelling-based Teaching in Science Education*. Cham: Springer International Publishing.
- Gogolin, S. & Krüger, D. (2016): Konstruktion von Diagnoseaufgaben zum Zweck von Modellen. In: *Biologie Lehren und Lernen – Zeitschrift für Didaktik der Biologie*, 20, 1, 44-62.
- Gogolin, S.; Krell, M.; Lange-Schubert, K.; Hartinger, A.; Upmeyer zu Belzen, A. & Krüger, D. (2017): Erfassung von Modellkompetenz bei Grundschüler/innen. In: H. Giest; A. Hartinger & S. Tänzler (Hrsg.): *Vielperspektivität im Sachunterricht*. Bad Heilbrunn: Klinkhardt, 108-115.
- Gogolin, S. (2017): *Diagnosing Students' Meta-Modelling Knowledge – Gathering Validity Evidence during Test Development*. Inaugural-Dissertation to obtain the academic degree

- Doctor rerum naturalium (Dr. rer. nat.), submitted to the Department of Biology, Chemistry and Pharmacy of Freie Universität Berlin.
- Gogolin, S. & Krüger, D. (2017): Diagnosing students' understanding of the nature of models. In: *Research in Science Education*, 47, 1127-1149.
- Gogolin, S. & Krüger, D. (2018): Students' understanding of the nature and purpose of models. In: *Journal of Research in Science Teaching*. URL: <https://doi.org/10.1002/tea.21453> [20.09.2018].
- Grosslight, L.; Unger, C. & Jay, E. (1991): Understanding Models and their Use in Science: Conceptions of Middle and High School Students and Experts. In: *Journal of Research in Science Teaching*, 28, 9, 799-822.
- Grünkorn, J.; Lotz, A. & Terzer, E. (2014): Erfassung von Modellkompetenz im Biologieunterricht. In: *Mathematischer und Naturwissenschaftlicher Unterricht (MNU)*, 67, 132-138.
- Hartig, J.; Klieme, E. & Leutner, D. (2008): Assessment of competencies in educational contexts: State of the art and future prospects. Göttingen: Hogrefe & Huber.
- Kane, M.T. (2013): Validating the Interpretations and Uses of Test Scores. In: *Journal of Educational Measurement*, 50, 1, 1-73.
- Koerber, S.; Sodian, B.; Kropf, N.; Mayer, D. & Schwippert, K. (2011): Die Entwicklung des wissenschaftlichen Denkens im Grundschulalter. Theorieverständnis, Experimentierstrategien, Dateninterpretation. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 1, 16-21.
- Koerber, S.; Mayer, D.; Osterhaus, C.; Schwippert, K. & Sodian, B. (2015): The development of scientific thinking in elementary school: A comprehensive inventory. In: *Child Development*, 86, 327-336.
- Köller, O. (2008): Lehr-Lern-Forschung. In: W. Schneider & M. Hasselhorn (Hrsg.): *Handbuch-Pädagogische Psychologie*. Göttingen: Hogrefe, 210-222.
- Krell, M.; Upmeyer zu Belzen, A. & Krüger, D. (2014): Context-Specificities in students' understanding of models and modelling. An issue of critical importance for both assessment and teaching. In: C. Constantinou; N. Papadouris & A. Hadjigeorgiou (Eds.): *E-Book Proceedings of the ESERA 2013 Conference*. Nicosia, 41-52. URL: <https://www.esera.org/publications/esera-conference-proceedings/esera-2013> [11.12.2017].
- Krell, M.; Upmeyer zu Belzen, A. & Krüger, D. (2012): Students' understanding of the purpose of models in different biological contexts. In: *International Journal of Biology Education*, 2, 1-34.
- Krell, M.; Upmeyer zu Belzen, A. & Krüger, D. (2014): Context-specificities in students' understanding of models and modelling: An issue of critical importance for both assessment and teaching. In: C. Constantinou; N. Papadouris & A. Hadjigeorgiou (Eds.): *E-Book proceedings of the ESERA 2013 conference*. Science education research for evidence-based teaching and coherence in learning. Part 6. Nature of science: History, philosophy and sociology of science. Nicosia, Cyprus: European Science Education Research Association. URL: <https://www.esera.org/publications/esera-conference-proceedings/esera-2013> [11.12.2017].
- Krell, M.; Upmeyer zu Belzen, A. & Krüger, D. (2016): Modellkompetenz im Biologieunterricht. In: A. Sandmann & P. Schmiemann (Hrsg.): *Biologiedidaktische Forschung: Schwerpunkte und Forschungsstände*. Band 1. Berlin: Logos, 83-102.
- Kuckartz, U. (Hrsg.) (2014): *Mixed Methods. Methodologie, Forschungsdesigns und Analyseverfahren*. Wiesbaden: Springer VS.
- Lange, K. & Hartinger, A. (2014): Modellierungskompetenz – Konzeptionierungen und Verortung im Sachunterricht. In: H.-J. Fischer; H. Giest & M. Peschel (Hrsg.): *Lernsituationen und Aufgabenkultur im Sachunterricht*. Bad Heilbrunn: Klinkhardt, 165-172.
- Mayring, P. (2010): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. (11. aktualisierte und überarbeitete Auflage). Weinheim und Basel: Beltz Verlag.

- Moosbrugger, H. & Kelava, A. (2012): Testtheorie und Fragebogenkonstruktion (2. aktualisierte und überarbeitete Auflage). Berlin, Heidelberg: Springer.
- Möller, K. (2015): Genetisches Lernen und Conceptual Change. In: J. Kahlert; M. Fölling-Albers; M. Götz; A. Hartinger; S. Miller & S. Wittkowske (Hrsg.): Handbuch Didaktik des Sachunterrichts. Bad Heilbrunn: Klinkhardt, 243-248.
- National Research Council [NRC] (Ed.) (2011): National Science Education Standards. Washington, DC: National Academic Press.
- NGSS Lead States (2013): Next Generation Science Standards: For states, by states. Washington: The National Academies Press.
- Nicolaou, C. & Constantinou, C. (2014): Assessment of the Modeling Competence. In: Educational Research Review, 13, 52-73.
- Rost, J. (2004): Lehrbuch Testtheorie – Testkonstruktion. (2. Aufl.). Göttingen: Hans Huber.
- Sandmann, A. (2014): Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In: D. Krüger; I. Parchmann & H. Schecker (Hrsg.): Methoden in der naturwissenschaftsdidaktischen Forschung. Heidelberg: Springer, 179-188.
- Schwarz, C.V.; Reiser, B.J.; Davis, E.A.; Kenyon, L.; Achér, A. & Fortus, D. (2009): Developing a learning progression for scientific modeling. Making scientific modeling accessible and meaningful for learners. In: Journal of Research in Science Teaching, 46, 6, 632-654.
- Schwarz, C.V.; Reiser, B.J.; Acher, A.; Kenyon, L. & Fortus, D. (2012): MoDeLS: Challenges in defining a learning progression for scientific modeling. In: A. Alonzo & A. Gotwals (Eds.): Learning Progressions in Science (LeaPS). Boston, MA: Sense Publishers, 101-137.
- Sodian, B.; Thoermer, C.; Kircher, E.; Grygier, P. & Günther, J. (2002): Vermittlung von Wissenschaftsverständnis in der Grundschule. In: Zeitschrift für Pädagogik, 45. Beiheft, 192-206.
- Sodian, B. (2004): Kognitive Entwicklung. In: C. Eggers; H. Fegert & F. Resch (Hrsg.): Psychiatrie und Psychotherapie des Kindes- und Jugendalters. Heidelberg: Springer.
- Strike, K.A. & Posner, G.J. (1992): A Revisionist Theory of Conceptual Change. In: R.A. Duschl & R.J. Hamilton (Eds.): Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice. Albany: State University of New York Press, 147-176.
- Toth, E.E.; Klahr, D. & Chen, Z. (2000): Bridging research and practice: A research-based classroom intervention for teaching experimentation skills to elementary school children. In: Cognition and Instruction, 18, 423-459.
- Trier, U. (2013): „Modelle sind künstlich“. In: Grundschule, 45, 12-14.
- Upmeyer zu Belzen, A. & Krüger, D. (2010): Modellkompetenz im Biologieunterricht. In: Zeitschrift für Didaktik der Naturwissenschaften, 16, 41-57.
- Vo, T.; Forbes, C.T.; Zangori, L. & Schwarz, C. (2015): Fostering 3rd-grade students' use of scientific models with the water cycle: Elementary teachers' conceptions and practices. In: International Journal of Science Education, 37, 15, 2411-2432.
- Willard, T. & Roseman, J.E. (2010): Probing students' ideas about models using standards-based assessment items. Paper presented at the Annual International Conference of the National Association of Research in Science Teaching (NARST), Philadelphia, PA. URL: <https://models.northwestern.edu/docs/Models%20Willard%20NARST%203-19-10.pdf> [11.12.2017].
- Wirtz, M. & Caspar, F. (2002): Beurteilertübereinstimmung und Beurteilerverlässlichkeit. Göttingen: Hogrefe.