

MODELLKOMPETENZ IM BIOLOGIEUNTERRICHT

Empirische Analyse von Modellkompetenz bei Schülerinnen und Schülern der Sekundarstufe I mit Aufgaben im offenen Antwortformat

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

am Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von
JULIANE GRÜNKORN

Berlin, 2014

Diese Arbeit entstand in der Zeit zwischen September 2008 und Mai 2014 unter der Leitung von Prof. Dr. Dirk Krüger in der Arbeitsgruppe für Didaktik der Biologie des Instituts Biologie im Fachbereich Biologie, Chemie und Pharmazie an der Freien Universität Berlin. Die Arbeit wurde selbstständig verfasst und alle Hilfsmittel wurden entsprechend aufgeführt.

Erstgutachter: Prof. Dr. Dirk Krüger, Freie Universität Berlin

Zweitgutachterin: Prof. Dr. Annette Upmeyer zu Belzen,
Humboldt-Universität zu Berlin

Tag der Disputation: 25. Juli 2014

Inhaltsverzeichnis

ABKÜRZUNGSVERZEICHNIS	III
ABBILDUNGSVERZEICHNIS	IV
TABELLENVERZEICHNIS	V
ZUSAMMENFASSUNG	1
ABSTRACT	5
1 EINLEITUNG.....	9
2 THEORETISCHER HINTERGRUND.....	13
2.1 Modellierung von Kompetenzen	14
2.1.1 Kompetenz.....	15
2.1.2 Kompetenzmodelle	17
2.1.3 Überprüfung von Kompetenzmodellen	20
2.2 Perspektiven auf Modelle	22
2.2.1 Wissenschaftstheoretische Sichtweisen von Modellen.....	24
2.2.2 Empirische Perspektiven auf Modelle	27
2.2.3 Strukturierungen der Vorstellungen von Modellen	34
2.3 Theoretische Struktur von Modellkompetenz	40
2.3.1 Graduierung in Niveaus.....	42
2.3.2 Dimensionierung in Teilkompetenzen	44
3 ZIELSETZUNGEN DER ARBEIT	49
4 OPERATIONALISIERUNG DES KOMPETENZMODELLS DER MODELLKOMPETENZ	53
4.1 Methode der Testkonstruktion	54
4.1.1 Theoretische Fundierung	54
4.1.2 Testkonzeption	55
4.1.3 Aufgabenformat und -auswertung.....	58
4.1.4 Aufgabenentwicklung.....	66
4.1.5 Aufgabenüberprüfung und -selektion.....	70
4.2 Ergebnisse der Operationalisierung	74
4.2.1 Beitrag 1: Design and test of open-ended tasks to evaluate a theoretical structure of model competence	74
4.2.2 Beitrag 2: Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells der Modellkompetenz.....	77
4.3 Zusammenfassende Diskussion.....	80

4.4	Fazit	84
5	ANALYSE UND BESCHREIBUNG VON MODELLKOMPETENZ.....	85
5.1	Untersuchungsdesign und Methodik	86
5.1.1	Beschreibung des Datenmaterials	87
5.1.2	Aufbereitung und qualitative Inhaltsanalyse des Datenmaterials.....	89
5.1.3	Statistische Analysen	91
5.2	Ergebnisse der Analyse von Modellkompetenz	100
5.2.1	Beitrag 3: Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework	100
5.2.2	Beschreibung von Besonderheiten in den Antwortstrukturen der Lernenden.....	104
5.2.3	Analyseergebnisse im Rahmen der IRT	111
5.3	Zusammenfassende Diskussion	116
5.3.1	Inhaltliche Überprüfung der theoretischen Struktur (Beitrag 3)	117
5.3.2	Besonderheiten in den Antwortstrukturen der Lernenden	120
5.3.3	Analysen im Rahmen der IRT	123
5.4	Fazit	128
6	DIDAKTISCHE ABLEITUNGEN UND ANREGUNGEN.....	131
6.1	Eigene Beiträge zu didaktischen Anregungen	132
6.1.1	Beitrag 4: Erfassung von Modellkompetenz im Biologieunterricht	133
6.1.2	Beitrag 5: Schwimmen wie Fische im Wasser – Untersuchung der Schleimschicht von Fischen.....	134
6.1.3	Beitrag 6: Bau und Funktion der Fischhaut.....	136
6.2	Zusammenfassende Implikationen für die Unterrichtspraxis	137
7	IMPLIKATIONEN FÜR DIE ZUKÜNFTIGE FORSCHUNG	145
	LITERATURVERZEICHNIS	149
	PUBLIKATIONEN DIESER ARBEIT.....	171
	Beitrag 1	172
	Beitrag 2	186
	Beitrag 3	206
	Beitrag 4	241
	Beitrag 5	249
	Beitrag 6	256
	DANKSAGUNG.....	263
	ANHANG.....	265

Abkürzungsverzeichnis

A	Teilkompetenz „Alternative Modelle“ des Kompetenzmodells
Ä	Teilkompetenz „Ändern von Modellen“ des Kompetenzmodells
BIBD	balanced incomplete block design
bik	BMBF-Projekt „Biologie im Kontext“
E	Bildungsstandards im Kompetenzbereich „Erkenntnisgewinnung“
EAP	Expected-a-posteriori-Methode
EM	Teilkompetenz „Eigenschaften von Modellen“ des Kompetenzmodells
FC-Items	Forced-Choice-Items
IRT	Item-Response-Theorie
Jgst.	Jahrgangsstufe
KMK	Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland
KTT	Klassische Testtheorie
M	Mittelwert
Max.	Maximum
MC-Items	Multiple-Choice-Items
Min.	Minimum
MMX-Design	Multi-Matrix-Design
N	Anzahl der Gesamtmenge
n	Anzahl der Teilmenge
PISA	Programme for International Student Assessment
PV	Plausible Value
SD	Standardabweichung
SUMS	Fragebogen „Students' Understanding of Models in Science“
T	Teilkompetenz „Testen von Modellen“ des Kompetenzmodells
TIMSS	Trends in International Mathematics and Science Study
TS1	Teilstudie 1 zur Operationalisierung
TS2	Teilstudie 2 zur Operationalisierung
Var	Varianz
VOMMS	Fragebogen „My Views of Models and Modelling in Science“
wMNSQ	weighted Mean-Square
Z	Teilkompetenz „Zweck von Modellen“ des Kompetenzmodells
δ	Item-Delta (Schwellenparameter)
κ	Cohens Kappa
τ	Thurstonian Threshold

Abbildungsverzeichnis

Abb. 1: Gegenstand der theoretischen Modellierung von Modellkompetenz (Grafik nach Meyer, 2012)	13
Abb. 2: Beziehung zwischen Kompetenz und Performanz (nach Meyer, 2012)	16
Abb. 3: Wissenschaftstheoretische Perspektiven auf Modelle (verändert nach Terzer, 2013)	26
Abb. 4: Das Forschungsdesign und die drei Ziele des Dissertationsprojekts (Grafik nach Meyer, 2012)	50
Abb. 5: Erstes Ziel der Forschungsarbeit: Operationalisierung des Kompetenzmodells der Modellkompetenz (Grafik nach Meyer, 2012)	53
Abb. 6: Auswertungsschritte der Teilstudien 1 und 2 zur Operationalisierung (vgl. Gläser-Zikuda, 2005; Gropengießer, 2005; Mayring, 2010).....	60
Abb. 7: Überblick über die Qualitätssicherungsschleifen (grau hinterlegt sind Schritte der Aufgabenüberprüfung und -selektion).....	71
Abb. 8: Zweites Ziel der Forschungsarbeit: Analyse und Beschreibung von Modellkompetenz (Grafik nach Meyer, 2012).....	85
Abb. 9: Methodische Vorgehensweise bei der Analyse von Modellkompetenz (vgl. Gläser-Zikuda, 2005; Gropengießer, 2005; Mayring, 2010).....	87
Abb. 10: Characteristic Curves by Score mit geordneten Schwellen ($\delta_1 = 0.01$; $\delta_2 = 0.35$; $\delta_3 = 0.89$) am Beispiel der Aufgabe A1	114
Abb. 11: Ausschnitt der Characteristic Curves by Score mit ungeordneten Schwellen ($\delta_1 = -0.32$; $\delta_2 = -1.57$) ¹⁰ am Beispiel der Aufgabe Ä3.....	114
Abb. 12: Wright Map der eindimensionalen Skalierung mit den Personenfähigkeiten (links) und den Thurstonian Thresholds (rechts)	116
Abb. 13: Drittes Ziel der Forschungsarbeit: Didaktische Ableitungen und Anregungen aus den Analysen (Grafik nach Meyer, 2012).....	131
Abb. 14: Strukturmodell zum möglichen Zusammenhang der didaktisch orientierten Perspektiven mit den Teilkompetenzen.....	146
Abb. 15: Strukturmodelle zum möglichen Zusammenhang des Parallelisierens mit den Teilkompetenzen Testen von Modellen und Ändern von Modellen (links) bzw. allen Teilkompetenzen (rechts)	147

Tabellenverzeichnis

Tab. 1:	Niveaus und Teilkompetenzen des theoriegeleiteten Kompetenzmodells der Modellkompetenz (verändert nach Upmeier zu Belzen & Krüger, 2010)	41
Tab. 2:	Strukturierung der aus der Literatur genannten Aspekte von Modellen zu Teilkompetenzen (verändert nach Krell, Upmeier zu Belzen & Krüger, 2014a sowie Upmeier zu Belzen & Krüger, 2010).....	46
Tab. 3:	Anleitung zur Aufgabenkonstruktion für die Teilkompetenz <i>Eigenschaften von Modellen</i>	67
Tab. 4:	Angaben zur Stichprobe der Studie zur Analyse von Modellkompetenz (vgl. Grünkorn, Upmeier zu Belzen & Krüger, 2014)	88
Tab. 5:	Multi-Matrix-Design – (15, 3, 1)-BIBD (Colbourn & Dinitz, 1996) zur Analyse von Modellkompetenz (vgl. Terzer, 2013)	89
Tab. 6:	Kennwerte zur Beurteilung des Leistungstests in einem probabilistischen Ansatz (Bühner, 2010; Hartig, 2009; Kelava & Moosbrugger, 2012)	96
Tab. 7:	Kategorien und ihre Häufigkeiten (%) ^a innerhalb der Niveaus sowie Gesamt-Häufigkeiten (%) ^b pro Niveau der Teilkompetenz (verändert nach Grünkorn, Upmeier zu Belzen & Krüger, 2014).....	102
Tab. 8:	Zusätzliche Kategorien und ihre Häufigkeiten (%) [*] zu den Teilkompetenzen (verändert nach Grünkorn, Upmeier zu Belzen & Krüger, 2014).....	103
Tab. 9:	Beschreibung der didaktisch orientierten Perspektiven mit Beispielaussagen und Häufigkeiten (%) [*] bezogen auf alle Teilkompetenzen.....	106
Tab. 10:	Die Häufigkeiten (%) [*] der didaktisch orientierten Kategorien in den fünf Teilkompetenzen des Kompetenzmodells.....	106
Tab. 11:	Kennwerte der eindimensionalen Raschskalierung auf Aufgabenebene	112
Tab. 12:	Kennwerte der eindimensionalen Raschskalierung auf Ebene der Antwortkategorien (0, 1, 2 und 3)	113
Tab. 13:	Mittelwerte der Personenfähigkeiten auf Ebene der Antwortkategorien (0, 1, 2 und 3).....	115
Tab. 14:	Aufgaben mit negativen punktbiserialen Korrelationen für die höchste Kategorie, mit ungeordneten Schwellenparametern und mit abnehmenden Mittelwerten der Personenfähigkeiten	124

Zusammenfassung

Die Kompetenz, Modelle als Mittel der Veranschaulichung und als Mittel der Erkenntnisgewinnung wahrzunehmen und einzusetzen, ist ein zentraler Bestandteil eines elaborierten Wissenschaftsverständnisses und der naturwissenschaftlichen Grundbildung (*scientific literacy*). Empirische Studien über Modelle und über den Modellbildungsprozess zeigen, dass Schülerinnen und Schüler Modelle vorwiegend als Mittel der Veranschaulichung verstehen und somit eine mediale Sichtweise auf Modelle einnehmen. Das Potenzial von Modellen, als Mediatoren zu fungieren, die zwischen Theorie und Phänomen vermitteln, damit sich neue Erkenntnisse über Phänomene gewinnen lassen (methodische Sichtweise), wird bislang nicht umfassend wahrgenommen.

Bei der Entwicklung einer medialen und methodischen Sichtweise auf Modelle helfen Kompetenzmodelle, die das entsprechende Konstrukt strukturieren und graduieren. Diese Kompetenzmodelle dienen Lehrkräften als Referenzrahmen für die Entwicklung von Fördermaßnahmen. Upmeyer zu Belzen und Krüger (2010) entwickelten ein theoretisch begründetes Kompetenzmodell der Modellkompetenz, das die empirischen Befunde und die wissenschaftstheoretische Literatur zu Modellen integriert. Es ist in fünf Teilkompetenzen – *Eigenschaften von Modellen*, *Alternative Modelle*, *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen* – strukturiert und graduieren diese Teilkompetenzen in jeweils drei Niveaus.

Eine empirische Überprüfung dieser Strukturierung und Graduierung steht jedoch aus. Auf der Basis einer solchen Überprüfung lässt sich beurteilen, inwiefern das Kompetenzmodell die Perspektiven von Lernenden angemessen reflektiert. Ergänzend dazu ermöglicht es Lehrkräften eine evidenzbasierte Entwicklung von Fördermaßnahmen. Dementsprechend stehen in der vorliegenden Arbeit die Operationalisierung des Kompetenzmodells der Modellkompetenz mit Aufgaben im offenen Antwortformat und, darauf aufbauend, die Analyse und Beschreibung der Modellkompetenz von Schülerinnen und Schülern der Sekundarstufe I im Mittelpunkt. Auf dieser Grundlage wurden didaktische Anregungen für einen (modell-)kompetenzorientierten Unterricht aus den empirischen Analysen abgeleitet.

Das Ziel, die Modellkompetenz der Lernenden zu analysieren und zu beschreiben, kann mit Aufgaben im offenen Antwortformat erreicht werden, deren

Antworten valide als Indikatoren von Modellkompetenz interpretiert werden können. Deshalb erfolgte eine systematische und theoriegeleitete Operationalisierung, welche die Kriterien der **Verständlichkeit und der Plausibilität des Aufgabenstamms**, des **Bezugs zur jeweiligen Teilkompetenz** und der **Repräsentation des Spektrums der jeweiligen Teilkompetenz** berücksichtigte.

Für die fünf Teilkompetenzen wurden insgesamt 40 Aufgaben entwickelt. Diese wurden zunächst mit sieben Expertinnen und Experten der Didaktik der Biologie nach den genannten Kriterien diskutiert und in zwei Teilstudien mit zusammen $N = 1\,231$ Schülerinnen und Schülern erprobt. Die Antworten und Kommentare der Lernenden wurden mit der qualitativen Inhaltsanalyse ausgewertet.

Darüber hinaus dienten beide Teilstudien dazu, Kodierleitfäden für die einzelnen Teilkompetenzen zu entwickeln, die neben Hinweisen zur Kodierung auch ein Kategoriensystem pro Teilkompetenz umfassen. Diese Kategorien fassen ähnliche Perspektiven der Lernenden auf Modelle und auf den Modellbildungsprozess zusammen.

Die Ergebnisse der Untersuchungen zeigten, dass insgesamt fünfzehn Aufgaben im offenen Antwortformat (drei Aufgaben pro Teilkompetenz) die oben genannten Kriterien erfüllen und daher als Indikatoren von Modellkompetenz interpretierbar sind. Diese Aufgaben nehmen die kognitive Komponente von Modellkompetenz in den Blick und eignen sich dazu, verschiedene Perspektiven von Schülerinnen und Schülern auf Modelle und auf den Modellbildungsprozess zu erfassen. Somit können diese Aufgaben zur Überprüfung der inhaltlichen Struktur des Kompetenzmodells herangezogen werden.

Für die Analyse und Beschreibung von Modellkompetenz wurde eine Querschnittsstudie mit $N = 1\,177$ Schülerinnen und Schülern der siebten bis zehnten Jahrgangsstufe durchgeführt. Die so gewonnenen Daten wurden ebenfalls mit der qualitativen Inhaltsanalyse ausgewertet, die Aufschluss über die Perspektiven der Lernenden auf Modelle und auf den Modellbildungsprozess gibt. Darauf aufbauend wurde geprüft, inwieweit das Kompetenzmodell der Modellkompetenz diese Perspektiven hinreichend beschreibt. Die Ergebnisse zeigten, dass für die Teilkompetenzen **Alternative Modelle** (nur ein Modell eines Originals), **Testen von Modellen** (Keine Testung des Modells) und **Ändern von Modellen** (Keine Änderung des Modells) empfohlen wird, ein zusätzliches basales Niveau zu berücksichtigen. Mit den Teilkompetenzen **Eigenschaften von Model-**

len und *Zweck von Modellen* hingegen werden die Perspektiven der Lernenden umfassend beschrieben.

Ergänzend dazu wurde anhand dieser Daten untersucht, ob sich bei den Lernenden dieser Stichprobe Besonderheiten in den Antwortstrukturen identifizieren ließen, um ein vertieftes Verständnis der Sichtweisen von Lernenden auf Modelle zu erhalten. Im Hinblick darauf konnten bei den Schülerinnen und Schülern dieser Stichprobe didaktisch orientierte Perspektiven (*Verständlichkeit/Kommunizierbarkeit* und *Zugänglichkeit*) identifiziert werden, die sich über alle fünf Teilkompetenzen und eine mediale und methodische Perspektive (bzw. Niveaus) erstrecken. Weiterhin zeigten die Befunde, dass bei Lernenden zwischen einem *abstrakten* und einem *konkreten Modellverständnis* unterschieden werden kann und dass das Parallelisieren von Modell und Original in den Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* zentral ist. In Bezug auf das Parallelisieren war auffällig, dass die Lernenden dies zum Teil ohne Berücksichtigung des Zwecks vornahmen. Gerade die Zweckgebundenheit spielt jedoch bei der Konstruktion von Modellen eine große Rolle und sollte beachtet werden.

Über diese Analysen hinaus wurde geprüft, inwieweit der entwickelte Leistungstest auch in einem probabilistischen Ansatz einsetzbar ist. Dies ist insbesondere im Hinblick auf Projekte zur längsschnittlichen Entwicklung von Modellkompetenz bei Schülerinnen und Schülern sinnvoll, die mit dem hier entwickelten Leistungstest einen solchen Ansatz verfolgen. Für die Prüfung wurde im Rahmen der Item-Response-Theorie ein ordinales, mehrdimensionales Rasch-Modell, und zwar das Partial-Credit-Modell, verwendet. Es ist festzuhalten, dass ein Einsatz in einem probabilistischen Ansatz möglich ist, wenn weitere Aufgaben entwickelt werden oder der Test mit anderen Tests zur Modellkompetenz kombiniert wird. Auf diese Weise lässt sich das Fähigkeitsspektrum optimal abdecken. Die Befunde der Analysen weisen zudem darauf hin, dass die Graduierung insbesondere hinsichtlich der Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* zu diskutieren ist.

Aus den Befunden zu den Perspektiven von Lernenden auf Modelle wurden didaktische Anregungen und Empfehlungen abgeleitet. Diese schließen sowohl Möglichkeiten zur Erfassung von Modellkompetenz im Unterricht anhand eines Kompetenzrasters und Diagnosebogens ein als auch zwei konkrete Unterrichtskonzepte, die Anregungen dazu geben, wie die Modellkompetenz im Bio-

logieunterricht bei Schülerinnen und Schüler der fünften und sechsten sowie der siebten und achten Jahrgangsstufe gefördert werden kann.

Schließlich werden mit Blick auf anschließende Forschungsvorhaben unter anderem weiterführende Fragen bezüglich der Dimensionalität des Kompetenzmodells, der Bedeutung didaktisch orientierter Kategorien (***Verständlichkeit/Kommunizierbarkeit*** und ***Zugänglichkeit***) und der Bedeutung des Parallelisierens von Modell und Original diskutiert.

Abstract

The competence of perceiving and implementing models as a means of illustration and gaining insights is a crucial element of an elaborated understanding of science and of *scientific literacy*. Empirical studies of models and the process of modelling have demonstrated that students predominantly perceive models as a means of illustration thus taking a medial perspective on models. So far, the potential of models regarding their function as mediators between theories and phenomena for the purpose of generating new insights into phenomena (methodical perspective) has not been properly explored.

The development of a medial and a methodical perspective on models is supported by competence models that structure and grade a respective construct. Teachers can use such competence models as a frame of reference for the development of interventions. Upmeyer zu Belzen and Krüger (2010) developed a theoretically founded competence model of model competence that integrates science-theoretical approaches and empirical findings on models. Their competence model is structured by five aspects: *nature of models, multiple models, purpose of model, testing models, and changing models; each of these* five aspects is subdivided into three levels.

This structuring and grading still lacks empirical evaluation, which would serve to assess whether the competence model adequately reflects the learners' perspectives. Furthermore, it would provide teachers with evidence that can be useful in developing intervention measures. Accordingly, the present study focuses on an operationalisation of the competence model by means of open-end items and subsequent analysis and description of model competencies of lower secondary school students. Based on the empirical analyses didactic implications for a model competence orientated instruction can be made.

The objective of analysing and describing the learners' model competencies can be achieved by presenting open-end items if it is possible to interpret responses as valid indicators of model competence. Therefore, the operationalisation was systematic and theory-based, taking into account the criteria of *comprehensibility and plausibility of items, reference to a respective aspect, and representation of the scope covered by an aspect (level I-III)*.

In total, 40 items were developed to assess the five aspects of model competence. Initially, the items were discussed by seven experts on model compe-

tence, according to the criteria listed above. They were subsequently tested with two partial studies involving a total number of $N = 1\,231$ students. Responses and comments from the learners were assessed using qualitative content analysis.

Based on both partial studies, coding guides were developed for the five aspects. Apart from providing coding information, the guides also contain a category system for each aspect. The categories comprise similar learner perspectives regarding models and the process of modelling.

Findings from the study showed that altogether 15 open-end items (three items per aspect) meet the criteria given above. Therefore, they can be interpreted as indicators of model competence. These items take the cognitive component of model competence into perspective and they are suitable for the assessment of different student perspectives on models and the process of modelling. It is therefore possible to apply these items to a test of the content structure of the competence model.

To analyse and describe model competence, a cross-sectional study was conducted involving $N = 1\,177$ students from grades seven to ten. Data from this assessment were also subjected to a qualitative content analysis making it possible to get insights into **students' perspectives** on models and the process of modelling. Based on this analysis, the research question on whether the competence model of model competence sufficiently and adequately describes these perspectives could be investigated. The results revealed that consideration of an additional basal level is recommended for the aspects of *alternative models* (only one model to one original), *testing models* (no testing of the model), and *changing models* (no change of the model). Regarding *nature of models* and *purpose of models*, the learners' perspectives are comprehensively described.

In addition, the data were analysed to find out whether learners from this sample showed particularities in response structure, to gain a deeper insight into learner perspectives on models. In this regard, didactically oriented perspectives could be identified in students from the sample, i.e. *intelligibility/communication* and *accessibility*, spanning all five aspects and medial and methodical perspectives (levels). The findings furthermore indicated that it is possible to distinguish between an *abstract* and a *concrete understanding of models* and that in the case of *testing models* and *changing models*, it is cru-

cial to compare the model with the original. It is evident that in some cases students draw comparison without considering the **model's purpose**. However, the construction of a model is always linked to a specific purpose.

Moreover, it was tested in how far the developed achievement test would also be implementable in a probabilistic approach. This is particularly useful regarding projects that focus on the longitudinal development of student model competencies and pursue such an approach based on the developed test. Based on Item-Response-Theory an ordinal multi-dimensional Rasch model was applied, i.e. the Partial-Credit Model. Accordingly, implementation in a probabilistic approach is possible if further items are developed or the test is combined with other tests assessing model competence. An optimal coverage of the scope of skills is thus possible. Furthermore, findings from the analyses indicate that grading is debatable particularly regarding the aspects of **testing models** and **changing models**.

Didactic recommendations and suggestions were deduced from the findings on learner perspectives on models. On the one hand, possibilities of assessing model competence in instruction by means of a competence grid and diagnostic sheet were addressed. On the other hand, two concrete lesson concepts were suggested offering incentives for fostering model competence in biology lessons in the fifth and sixth form respectively seventh and eighth form.

With regard to further research projects, the following aspects are also discussed: further questions about the dimensionality of the competence model, the relevance of didactically oriented categories (**intelligibility/communication and accessibility**), and the meaning of comparing a model with an original.

1 Einleitung

Mit dem Perspektivenwechsel von einem input- hin zu einem outcome-orientierten Denken im deutschen Bildungssystem rückt die Vermittlung von Kompetenzen, das heißt von relevanten Kenntnissen und Fertigkeiten in konkreten Anforderungsbereichen, verstärkt in den Fokus von Schulen und Öffentlichkeit (Klieme, Avenarius et al., 2007; Klieme, Hartig & Rauch, 2008; Köller, 2009). In den Naturwissenschaften spielt die Kompetenz, in Modellen zu denken und mit Modellen reflektiert umzugehen, eine wichtige Rolle. So nutzen Forschende Modelle als Instrumente zur Generierung neuen Wissens und für den wissenschaftlichen Austausch (Clement, 2000; Coll, France & Taylor, 2005; Oh & Oh, 2011). Ein entscheidender Teil der Tätigkeiten von Forschenden liegt darin, Modelle zu entwickeln und mit ihnen Hypothesen über Phänomene zu überprüfen bzw. neue Hypothesen abzuleiten (Giere, 2004; Greca & Moreira, 2000; Mahr, 2009; Odenbaugh, 2005; Oh & Oh, 2011).

Nicht nur für den wissenschaftlichen Erkenntnisprozess kommt Modellen eine zentrale Bedeutung zu. Auch im beruflichen und privaten Alltag von Bürgerinnen und Bürgern sind sie allgegenwärtig. So fertigen Försterinnen und Förster Zonierungsmodelle von ihren Wäldern an, um Veränderungsprozesse kenntlich zu machen und Voraussagen über die Entwicklung ihres Waldes zu tätigen. Auch in der medialen Berichterstattung wird die Fähigkeit, in Modellen zu denken, (implizit) vorausgesetzt, beispielsweise wenn Bürgerinnen und Bürger mit verschiedenen Modellen zur gesunden Ernährung, zum Beispiel Ernährungspyramiden, konfrontiert werden. Um praktische Lebensanforderungen bewältigen und aktiv am gesellschaftlichen Leben teilnehmen zu können, ist daher ein kompetenter Umgang mit Modellen unerlässlich. Deshalb ist es wichtig, Wissen über die Rolle von Modellen für den wissenschaftlichen Erkenntnisprozess zu erwerben und darüber zu reflektieren (Grosslight, Unger, Jay & Smith, 1991).

Aufgrund dieser Relevanz von Modellen legen auch nationale und internationale naturwissenschaftliche Bildungsstandards normativ einen Schwerpunkt darauf, was Schülerinnen und Schüler über Modelle und den Modellbildungsprozess wissen sollen und wie sie damit umgehen können sollen (National Research Council, 2012; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005). Demnach sollen Schülerinnen und Schüler in der Lage sein, mit Modellen komplexe wissen-

schaftliche Konzepte anschaulich zu erklären (*mediale Perspektive*; Gilbert, 1991; Mahr, 2008; Oh & Oh, 2011). Zudem sollen sie im Unterricht Modelle bauen und anwenden sowie über ihre Aussagekraft reflektieren (*methodische Perspektive*; Gilbert, 1991; Mahr, 2008; Oh & Oh, 2011). Die Berücksichtigung dieser medialen und methodischen Perspektive auf Modelle im Unterricht trägt dazu bei, eine umfassende Modellkompetenz bei Lernenden zu entwickeln. Darüber hinaus können mit Modellen sowohl Fachwissen als auch Charakteristika der Naturwissenschaften und ihrer Methoden vermittelt werden (Clement, 2000; Gilbert & Boulter, 2000; Henze, van Driel & Verloop, 2007; Hodson, 1993; Schwarz & White, 2005).

Empirische Studien zeigen, dass Lernende Modelle vor allem unter der medialen Perspektive betrachten und ihre Rolle im wissenschaftlichen Prozess der Erkenntnisgewinnung kaum wahrnehmen (z. B. Fleige, Seegers, Upmeier zu Belzen & Krüger, 2012b; Grosslight et al., 1991; Grünkorn, Upmeier zu Belzen & Krüger, 2014; Harrison & Treagust, 2000; Krell, Upmeier zu Belzen & Krüger, 2012b; Orsenne & Upmeier zu Belzen, 2012; Terzer, 2013; Trier & Upmeier zu Belzen, 2009). Hier besteht demzufolge ein Förderungsbedarf, um die Modellkompetenz von Lernenden auch im Sinne der Bildungsstandards (KMK, 2005) umfassend zu entwickeln.

Ausgehend von diesem Förderungsbedarf und der Bedeutung von Modellen für die Entwicklung eines elaborierten Wissenschaftsverständnisses (Clement, 2000; Gilbert & Boulter, 2000; Harrison & Treagust, 2000) hat man sich in Studien im schulischen Kontext bislang darauf konzentriert, verschiedene Aspekte von Modellen und der Modellbildung zu beschreiben (u. a. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Meisert, 2008; Oh & Oh, 2011; Schwarz et al., 2009). Diese Strukturierungsansätze unterscheiden sich unter anderem hinsichtlich der Dimensionierung, also der Einteilung in Aspekte, und der Graduierung, also der Abstufung in verschiedene Niveaus. Zudem fehlt eine umfassende empirische Überprüfung der meisten dieser Ansätze mit Lernenden unterschiedlichen Alters, die zeigt, inwiefern diese Aspekte die Perspektiven von Lernenden hinreichend berücksichtigen (vgl. Schwarz et al., 2009, S. 637).

Relevante Ergebnisse, die diesen Strukturierungsansätzen zugrunde liegen, integriert ein Kompetenzmodell der Modellkompetenz von Upmeier zu Belzen und Krüger (2010). Dieses Kompetenzmodell basiert sowohl auf empirischen

Studien (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003) als auch auf wissenschaftstheoretischer Literatur zu Modellen (u. a. Gilbert, 1991; Mahr, 2008, 2009). Es differenziert fünf Teilkompetenzen: **Eigenschaften von Modellen**, **Alternative Modelle**, **Zweck von Modellen**, **Testen von Modellen** und **Ändern von Modellen** sowie drei Reflexionsniveaus für jede Teilkompetenz. Mit der theoriegeleiteten Formulierung kommen Upmeier zu Belzen und Krüger (2010) der Forderung nach, die Entwicklung von Kompetenzmodellen theoretisch zu begründen, so dass Modellkompetenz auf der Grundlage ihres Kompetenzmodells operationalisiert werden kann und sich entsprechende Messinstrumente entwickeln lassen (Klieme, Avenarius et al., 2007; Pellegrino, Chudowsky & Glaser, 2001).

Erst eine empirische Überprüfung, in der geprüft wird, inwiefern diese Strukturierung die Perspektiven von Lernenden angemessen reflektiert (vgl. Klieme, Avenarius et al., 2007), kann Lehrkräften Evidenzen für die Entwicklung von Lehrplänen und Fördermaßnahmen bieten (Klieme, Maag Merki & Hartig, 2007). Die vorliegende Arbeit leistet hierzu einen Beitrag. In einem Kooperationsprojekt zur Modellkompetenz zwischen der Humboldt-Universität zu Berlin und der Freien Universität Berlin realisierten Terzer (2013) die empirische Überprüfung des Kompetenzmodells mit Multiple-Choice-Items (MC-Items) und Krell (2013) mit Forced-Choice-Items (FC-Items). Die hier beschriebene Arbeit operationalisiert die kognitive Komponente von Modellkompetenz mit Aufgaben im offenen Antwortformat (Grünkorn & Krüger, 2012; Grünkorn, Upmeier zu Belzen & Krüger, 2011). Damit ist es möglich, verschiedene Perspektiven von Lernenden der Sekundarstufe I auf Modelle zu identifizieren. Diese wiederum erlauben es, das Kompetenzmodell inhaltlich zu überprüfen und die einzelnen Niveaus der jeweiligen Teilkompetenz ausdifferenzieren (Grünkorn, Upmeier zu Belzen et al., 2014). Aufbauend darauf lassen sich didaktische Anregungen für Lehrkräfte ableiten (Grünkorn & Fleige, 2012; Grünkorn & Hanauer, 2013a, 2013b; Grünkorn, Lotz & Terzer, 2014), welche die Lehrkräfte dabei unterstützen können, den Herausforderungen eines (modell)kompetenzorientierten Unterrichts zu begegnen (vgl. KMK, 2010).

2 Theoretischer Hintergrund

Theoretischer Schwerpunkt dieser Forschungsarbeit ist das Konstrukt *Modellkompetenz*. Dieses Konstrukt ist latent, also nicht direkt beobachtbar, weshalb es zunächst theoretisch modelliert und beschrieben wird (vgl. Klieme et al., 2008; Köller, 2008; Abb. 1). Mit dem Verständnis von Modellkompetenz als latente Fähigkeit sind zunächst die Kompetenzdefinition (vgl. Klieme & Leutner, 2006; Weinert, 2001) und die Modellierung von Kompetenzen in Form von Kompetenzmodellen die theoretische Grundlage dieser Arbeit. Hinsichtlich der Frage nach einer adäquaten Vorgehensweise für eine empirische Überprüfung dienen andere Projekte als Orientierungshilfe, die ebenfalls Kompetenzmodelle mit offenen Formaten untersuchten (z. B. Eggert & Bögeholz, 2010; Grube, 2011; vgl. S. 14ff. dieser Arbeit).

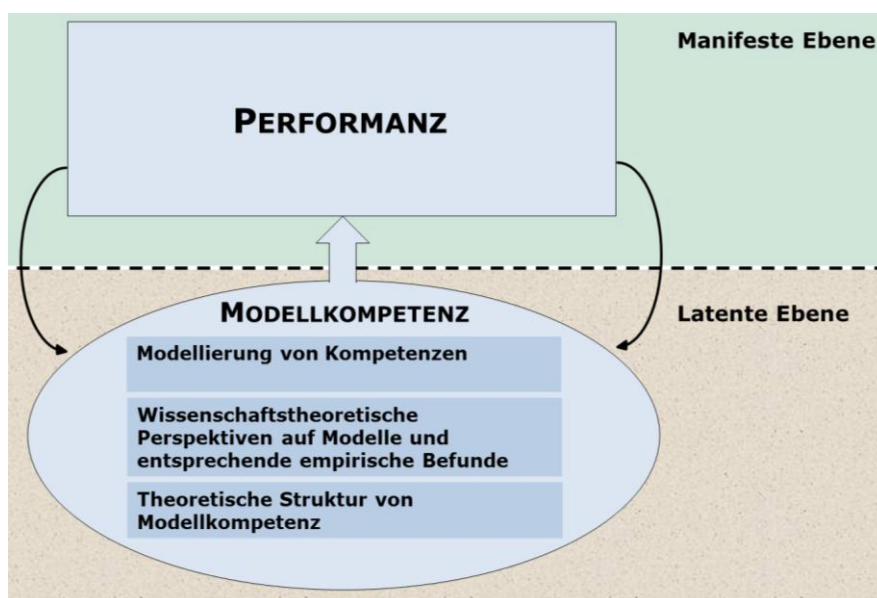


Abb. 1: Gegenstand der theoretischen Modellierung von *Modellkompetenz* (Grafik nach Meyer, 2012)

Ein weiterer zentraler Gesichtspunkt der theoretischen Modellierung von Modellkompetenz ist neben der Modellierung von Kompetenzen insbesondere die wissenschaftstheoretische Perspektive auf Modelle als Medien und als Methode (z. B. Mahr, 2008; Stachowiak, 1973). Dem stehen unterschiedliche empirische Befunde zu den Perspektiven von Lernenden auf Modelle und auf den Modellbildungsprozess gegenüber, die von Forschenden in unterschiedliche Aspekte strukturiert und graduiert sind (z. B. Crawford & Cullin, 2005; Grosslight

et al., 1991; Justi & Gilbert, 2003; vgl. S. 22ff. dieser Arbeit). Aufbauend darauf haben Upmeyer zu Belzen und Krüger (2010) eine theoretische Struktur von Modellkompetenz entwickelt, die die Grundlage dieser Forschungsarbeit darstellt (vgl. S. 40ff. dieser Arbeit).

Auf der Basis der theoretischen Modellierung kann Modellkompetenz mit Aufgaben operationalisiert und in Antworten von Lernenden erfasst werden. Von der manifesten Performanz lässt sich dann auf die latente Kompetenz der Lernenden schließen (vgl. Klieme et al., 2008; Köller, 2008; Abb. 1). Zudem erlaubt die Erfassung der Performanz die empirische Überprüfung der theoretischen Modellierung des Konstrukts.

2.1 Modellierung von Kompetenzen

Das Wort *Kompetenz* wird in der deutschen Alltagssprache häufig und mit unterschiedlicher Bedeutung benutzt (Klieme & Hartig, 2007). So titeln Zeitungen, beispielsweise *Die Welt*, im Bereich Politik „Cameron will Kompetenzen von EU zurückholen“ (Die Welt, 2013) oder in der Sportberichterstattung „Im Aufsichtsrat fehlt die sportliche Kompetenz“ (Laux, 2013). Ersteres bezieht sich auf Zuständigkeiten oder auf Handlungsbefugnisse, letzteres beschreibt dagegen die Befähigung von Personen, spezifische Anforderungen zu bewältigen (Schott & Azizi Ghanbari, 2008). Demnach hat „Kompetenz [...] offenbar irgendwie zu tun mit Zuständigkeit und mit Fähigkeit und mit Bereitschaft und damit, daß Zuständigkeit, Fähigkeit und Bereitschaft sich in Deckung befinden“ (Marquard, 1981, S. 24).

Auch in den verschiedenen an der Bildungsforschung beteiligten Disziplinen ist der Kompetenzbegriff allgegenwärtig und hat vielfältige Bedeutungen. Das liegt vor allem an den verschiedenen traditionellen Ansätzen,¹ wie dem *generativen Ansatz* von Noam Chomsky, dem *normativen Ansatz* in den Erziehungswissenschaften und dem *funktional-pragmatischen Ansatz* von David McClelland und anderen in der Psychologie (vgl. Hartig, 2007; Klieme et al., 2008). Diese Bedeutungsvielfalt hat zur Folge, dass Missverständnisse entstehen und Studienergebnisse nur begrenzt vergleichbar sind (Hartig, 2008; Klieme, Maag Merki et al., 2007). Aus diesem Grund wird der dieser Arbeit zu-

¹ Für eine ausführliche Beschreibung der verschiedenen traditionellen Ansätze siehe Klieme und Hartig (2007) sowie Klieme, Hartig und Rauch (2008).

grunde liegende Kompetenzbegriff beschrieben und von anderen Konstrukten abgegrenzt (vgl. S. 15ff. dieser Arbeit). Kompetenzkonstrukte werden in Kompetenzmodellen strukturiert und beschrieben (Klieme & Hartig, 2007; Klieme, Maag Merki et al., 2007). Diese sind idealerweise theoriegeleitet entwickelt, um curriculare Anforderungen, fachspezifische Zusammenhänge und kognitionspsychologische Theorien einzubinden (Klieme & Leutner, 2006; vgl. S. 17ff. dieser Arbeit). Neben der theoretischen Fundierung ist eine empirische Überprüfung von Kompetenzmodellen notwendig, um zu ermitteln, inwiefern die postulierten Strukturen die jeweiligen Perspektiven von Personen hinreichend widerspiegeln (Klieme, Avenarius et al., 2007; Schecker & Parchmann, 2006). Im Fach Biologie ist das Projekt *Biologie im Kontext (bik)* (Bayrhuber, Bögeholz, Eggert et al., 2007) ein Wegbereiter auf dem Feld der empirischen Überprüfung von Kompetenzmodellen gewesen. In diesem Projekt wurden Kompetenzmodelle für verschiedene Kompetenzbereiche theoretisch entwickelt und auf unterschiedliche Weise empirisch überprüft (Bayrhuber, Bögeholz, Eggert et al., 2007; Bayrhuber, Bögeholz, Elster et al., 2007). An diesen Ansätzen orientierte sich die hier beschriebene empirische Überprüfung des Kompetenzmodells der Modellkompetenz (vgl. S. 20ff. dieser Arbeit).

2.1.1 Kompetenz

Der Kompetenzbegriff des funktional-pragmatischen Ansatzes, der Anfang der 1970er-Jahre in der Psychologie entwickelt wurde, fokussiert nicht auf allgemeine generative kognitive Systeme (*generativer Ansatz* von Noam Chomsky) oder auf allgemeine Bildungserwartungen (*normativer Ansatz* in den Erziehungswissenschaften), sondern auf die individuellen Fähigkeiten von Personen, die ihnen gestellten Anforderungen in spezifischen Situationen zu bewältigen (Klieme & Hartig, 2007; Klieme et al., 2008). Damit kommt dieser Ansatz der Forderung nach „testing for competence rather than for ,intelligence“ von McClelland (1973, S. 1) nach. Mit der Messung von Kompetenzen lassen sich somit Leistungsunterschiede in realen Situationen besser vorhersagen (Klieme, Maag Merki et al., 2007).

Ein Kernaspekt dieses Ansatzes ist die *Kontextspezifität* bzw. der Bezug zum „wirklichen Leben“ (Klieme, Maag Merki et al., 2007, S. 6). Connell, Scheridan und Gardner (2003) beschreiben Kompetenzen daher als „realized abilities“ (S. 142). Personen erwerben Kompetenzen danach in verschiedenen Kontex-

ten bzw. Situationen und können diese in ähnlichen Situationen wieder aufrufen und anwenden. Somit sind Kompetenzen funktional auf Situationen bezogen und durch Lernprozesse *erlernbar* und *beeinflussbar* (Baumert, Stanat & Demmrich, 2001; Simonton, 2003).

Nach dem funktional-pragmatischen Ansatz sind Kompetenzen darüber hinaus *Dispositionen*. Das bedeutet, sie sind nicht direkt beobachtbar und müssen daher theoretisch modelliert (Abb. 2, Tiefenstruktur) und mit Aufgaben operationalisiert (Abb. 2, Oberflächenstruktur) werden, um sie erfassen zu können. Dabei lassen sich direkt beobachtbare Leistungen (Performanz) im Sinne von Reaktionen auf Items von verfügbaren Tiefenstrukturen (latenten Fähigkeiten) unterscheiden (Köller, 2008; Meyer, 2012). Aus der in einem Test gezeigten Leistung einer Person kann auf die jeweilige Kompetenz geschlossen bzw. diese rekonstruiert werden (Klieme & Hartig, 2007, S. 27). Die Beziehung zwischen Kompetenz und Performanz lässt sich auf Pflanzen übertragen und daran – modellhaft – verdeutlichen (Abb. 2). Die Wurzel einer Pflanze, die sich unter der Erde befindet, repräsentiert die Tiefenstruktur (Kompetenz), die nicht direkt zu sehen ist. Diese Wurzel beeinflusst unter anderem das Wachstum (Performanz) der Sprossachse. Die Sprossachse repräsentiert die Oberflächenstruktur (Performanz), die von der beobachtenden Person wahrgenommen wird. Von der Qualität des Wachstums lassen sich nun Rückschlüsse auf die Wurzel ziehen.

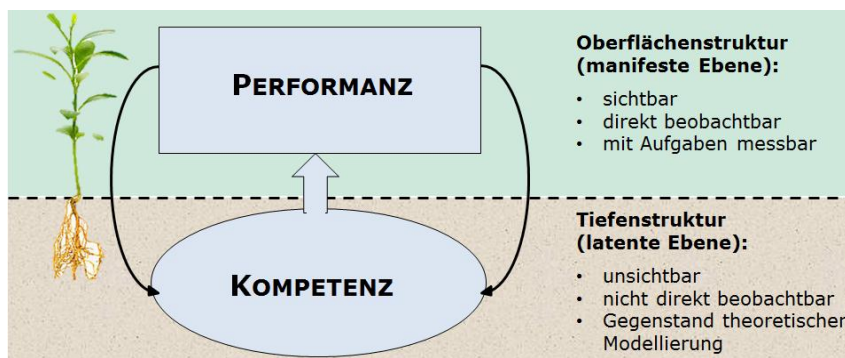


Abb. 2: Beziehung zwischen Kompetenz und Performanz (nach Meyer, 2012)

Zusammenfassend ist festzuhalten, dass Kompetenzen kontextspezifische Dispositionen sind, die sich funktional auf Situationen und spezifische curriculare Domänen beziehen und erlernbar sind. Damit grenzt sich das Kompetenzkonstrukt von anderen Dispositionen wie dem Intelligenzkonstrukt ab, das allge-

meine kontextunabhängige Bereiche beschreibt, wobei Intelligenz nur sehr begrenzt erlernbar ist (Klieme et al., 2008).

Nach diesem Kompetenzverständnis unterscheiden sich die Kompetenzdefinitionen vor allem hinsichtlich der Breite der Domäne – spezifische Kompetenzen oder breit angelegte Schlüsselkompetenzen – **und danach, welche „mentalen Bedingungen“** (Klieme & Hartig, 2007, S. 19) – unter anderem kognitive, motivationale, soziale, volitionale Komponenten – der Performanz zugrunde liegen (Klieme et al., 2008). So bezieht Weinert (1999, 2001) in seine Definition neben kognitiven Komponenten auch die motivationale, die volitionale und die soziale Bereitschaft ein. Zugleich empfiehlt er, kognitive Kompetenzen und nicht-kognitive Komponenten getrennt zu erfassen, um Zusammenhänge zwischen diesen Komponenten identifizieren zu können (Weinert, 1999, 2001). Diese Trennung, wie auch die Fokussierung auf kognitive Kompetenzen entsprechen den Kompetenzkonzepten der internationalen Schulleistungsstudien PISA (Programme for International Student Assessment) und TIMSS (Trends in International Mathematics and Science Study), denen der KMK-Bildungsstandards (Pant, 2013; Pant, Böhme & Köller, 2012) sowie der von Klieme und Leutner (2006) vorgelegten Definition. Diese beschreiben **Kompetenzen „als kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen“** (Klieme & Leutner, 2006, S. 4; vgl. Weinert, 1999, 2001). Basierend auf der Empfehlung von Weinert (1999, 2001), stehen die kognitiven Kompetenzen im Fokus der vorliegenden Arbeit. Die von Klieme und Leutner (2006) aufgeführte Begriffsbestimmung wird als Arbeitsdefinition herangezogen.

2.1.2 Kompetenzmodelle

In der Modellierung von Kompetenzen wird in der Literatur zwischen *Kompetenzstrukturmodellen*, *Kompetenzniveaumodellen* und *Kompetenzentwicklungsmodellen* unterschieden (Hartig & Klieme, 2006; Klieme, Maag Merki et al., 2007). Diese Arten der Modellierung schließen einander nicht aus, sondern ergänzen sich und können gegebenenfalls voneinander abgeleitet werden (z. B. ein Kompetenzentwicklungsmodell von einem Kompetenzniveaumodell; Klieme, Maag Merki et al., 2007; Robitzsch, 2013).

Kompetenzstrukturmodelle beschreiben, in welche und in wie viele Teilkompetenzen (*Dimensionen*², Hartig & Klieme, 2006) ein Kompetenzkonstrukt strukturiert ist und wie diese Teilkompetenzen miteinander zusammenhängen (Hartig & Klieme, 2006; Klieme, Maag Merki et al., 2007). Teilkompetenzen **beschreiben unterschiedliche Fähigkeiten und damit „verschiedene Arten von individuellen Ressourcen“ (Klieme, Maag Merki et al., 2007, S. 12). Diese „individuellen Ressourcen“ sind notwendig, damit Personen die ihnen gestellten Anforderungen in spezifischen Situationen erfolgreich bewältigen können (Klieme, Maag Merki et al., 2007).** Die Generierung der Teilkompetenzen erfolgt entweder im Nachhinein mittels faktorenanalytischer Methoden oder im Voraus mithilfe von theoretischen Annahmen, die zum Beispiel mit linearen Strukturgleichungsmodellen empirisch überprüft werden (Fleischer, Koeppen, Kenk, Klieme & Leutner, 2013). Beispiele für Kompetenzstrukturmodelle in den Naturwissenschaften sind das Modell der Kommunikationskompetenz (Kramer, 2009) oder das Modell des Umgangs mit Diagrammen (Lachmayer, 2008).

Kompetenzniveaumodelle befassen sich mit unterschiedlichen Ausprägungen innerhalb einer Teilkompetenz, die als Niveaus bezeichnet werden (Hartig, 2007; Hartig & Klieme, 2006). Im Modell von Bybee (1997, zitiert nach Schecker & Parchmann, 2006) beispielsweise werden vier Niveaus naturwissenschaftlicher Grundbildung (*scientific literacy*) unterschieden: (1) nominale, (2) funktionale, (3) konzeptionelle und prozedurale sowie (4) multidimensionale. Aufbauend darauf sind die fünf (Prenzel, Drechsel, Carstensen & Ramm, 2004) und später sechs (Prenzel et al., 2008) Stufen naturwissenschaftlicher Kompetenz bei PISA entstanden. Dabei wurden die **funktionalen** sowie **konzeptionellen und prozeduralen** Kompetenzen in jeweils zwei Stufen unterteilt. Bei PISA 2000 und 2003 wurde zudem die **multidimensionale** Kompetenz ausgeklammert und bei PISA 2006 wieder einbezogen. Erst mit der Beschreibung von Niveaus ist es möglich, inhaltlich zurückzumelden, über welche Kompetenzen Lernende verfügen oder nicht (kriteriumsorientierte Interpretation).

² Der Terminus *Dimension* findet auf methodischer als auch theoretischer Ebene unterschiedliche Verwendung. Mit Bezug auf psychometrische Modelle sind unter Dimensionen zusammenhängende Itemgruppen zu verstehen. Auf der Ebene der Kompetenzmodellierung werden Teilkompetenzen (z. B. Schmiemann, 2010) bzw. Ausprägungen innerhalb einer Teilkompetenz (z. B. Senkbeil, Rost, Carstensen & Walter, 2005) als Dimensionen bezeichnet oder mehrere Teilkompetenzen unter einer Dimensionen subsumiert (z. B. Schecker & Parchmann, 2006). In dieser Arbeit wird der Begriff *Dimension* in Bezug auf psychometrische Modelle benutzt und darunter werden zusammenhängende Itemgruppen verstanden. Auf theoretischer Ebene werden Dimensionen als übergeordnete Strukturelemente aufgefasst, die in verschiedene Teilkompetenzen untergliedert sind.

Bei der Rückmeldung von Kompetenzen herrschte bislang in der psychologischen Leistungsforschung eine normorientierte Interpretation (z. B. Intelligenzquotient-Norm) der Testwerte vor. Diese erlaubt es lediglich, in Bezug auf eine Vergleichsgruppe von hohen, niedrigen, durchschnittlichen oder überdurchschnittlichen Ausprägungen zu sprechen (Hartig & Klieme, 2006). In Schulleistungsstudien ist die normorientierte Interpretation allerdings eher unbefriedigend und wenig aussagekräftig, weshalb bei diesen Studien eine kriteriumsorientierte Interpretation der Testwerte bevorzugt wird (z. B. Helmke & Hosenfeld, 2004; Klauer, 1987). Die Niveaus für eine kriteriumsorientierte Interpretation sind entweder mittels nachträglicher Analysen oder bereits im Voraus modelliert (vgl. Cizek & Bunch, 2007). Letzteres stellt bislang noch eine Ausnahme dar, ist allerdings bei der Entwicklung von Kompetenzmodellen zu empfehlen, weil im Vorhinein formulierte Niveaus eine Operationalisierung und die damit verbundene Aufgabenentwicklung vereinfachen (Klieme & Leutner, 2006).

Niveaus beschreiben aus psychometrischer Sicht bestimmte Abschnitte auf einer Skala (Hartig, 2007). Diese Abschnitte können qualitativ unterschiedliche Ausprägungen (nominale Skala) oder hierarchische Ausprägungen (ordinale Skala) mit zunehmender Komplexität darstellen (Einhaus, 2007; Fleischer et al., 2013). Spiegeln die Abschnitte darüber hinaus eine zeitliche Entwicklung wider, berücksichtigen sie entwicklungspsychologische Voraussetzungen, Lernschwierigkeiten sowie vorunterrichtliche Vorstellungen und sind sie über eine Längsschnittstudie erforscht, wird bei diesen Abschnitten von Stufen und insgesamt von **Kompetenzentwicklungsmodellen** gesprochen (z. B. Hammann, 2004). Solche Kompetenzmodelle spielen insbesondere bei der unterrichtlichen Förderung von Kompetenzen eine wichtige Rolle, liegen aber nur selten vor (Fleischer et al., 2013; Hammann, 2004; Klieme et al., 2008; Schecker & Parchmann 2006). Im naturwissenschaftlichen Bereich steht aber zunehmend die Untersuchung der Entwicklungsaspekte von Kompetenzen im Vordergrund (z. B. Eggert, Bögeholz, Watermann & Hasselhorn, 2010; Grube, 2011).

Neben der Unterscheidung in **Kompetenzstrukturmodelle**, **Kompetenzniveau-modelle** und **Kompetenzentwicklungsmodelle** unterscheiden Schecker und Parchmann (2006) zwischen **normativen** und **deskriptiven** Kompetenzmodellen. Diese Differenzierung bezieht sich auf unterschiedliche Fundierungen. **Normative Kompetenzmodelle** (z. B. Bildungsstandards für das Fach Biologie, KMK,

2005) basieren häufig auf Bildungszielen oder leiten sich aus Theorien über kognitive Strukturen ab. Sie sind historisch gewachsen und beschreiben Kompetenzen, über welche Lernende verfügen sollen. Somit sind diese Kompetenzmodelle hypothetisch, weswegen eine empirische Überprüfung ihrer Struktur und der damit verbundenen Leistungserwartungen notwendig ist (Schecker & Parchmann, 2006).

Deskriptive Kompetenzmodelle (z. B. TIMSS: Klieme, Baumert, Köller & Bos, 2000; PISA: Adams & Wu, 2002) beschreiben hingegen typische Muster kognitiver Voraussetzungen, die in empirischen Untersuchungen identifiziert wurden (Schecker & Parchmann, 2006). Wenn beispielsweise mehrere Items ähnlich gelöst werden, könnte dieser Lösung eine gemeinsame Dimension zugrunde liegen. Diese Zusammenhänge bzw. Muster werden auf der Grundlage von Testanalyseverfahren (z. B. explorative Faktorenanalyse, Item-Response-Theorie-Analysen) formuliert (Einhaus, 2007; Schecker & Parchmann, 2006).

2.1.3 Überprüfung von Kompetenzmodellen

Klieme, Avenarius et al. (2007, S. 82) konstatieren in ihrer Expertise zur Entwicklung nationaler Bildungsstandards, dass **„empirische Untersuchungen mithilfe von Tests [...] erforderlich [sind], um zu prüfen, ob diese Modelle tatsächlich die Aspekte der Kompetenzen von Lernenden, ihre Niveaustufung und gegebenenfalls ihre Entwicklung angemessen widerspiegeln“**. Erst mit einer empirischen Überprüfung sind die Voraussetzungen dafür geschaffen, faire und valide Testinstrumente für Systemmonitoring, Schulevaluation und Individualdiagnostik einzusetzen.

Mit der Wahl eines solchen Einsatzbereichs eines Kompetenzmodells – Überprüfung von Kompetenzmodellen, Systemmonitoring, Schulevaluation oder Individualdiagnostik – sind an die Test- und Aufgabenkonstruktion sowie die Testorganisation und Testauswertung Bedingungen und Einschränkungen geknüpft (Kauertz, 2014; Klieme, Avenarius et al., 2007; Leutner, Fleischer, Spoden & Wirth, 2007; Neuhaus & Braun, 2007). Die Güte eines Tests ist somit mit Blick auf die Nutzung und die Art der Schlussfolgerung zu beurteilen (Klieme, Avenarius et al., 2007). Hinsichtlich der empirischen Überprüfung von Kompetenzmodellen bedeutet dies beispielsweise, dass das Konstrukt zwar ausreichend durch Aufgaben repräsentiert sein muss, dass allerdings für Tests auf Populationsebene auch geringere Messgenauigkeiten (Reliabilität) als für

Tests zur Individualdiagnostik akzeptabel sind (vgl. Terzer, Hartig & Upmeyer zu Belzen, 2013).

Für das Fach Biologie liegen sowohl theoretisch als auch empirisch fundierte Kompetenzmodelle vor. Im Projekt *bik* (Bayrhuber, Bögeholz, Eggert et al., 2007) haben Forschungsgruppen Kompetenzmodelle theoretisch hergeleitet und empirisch überprüft, um Kompetenzen von Schülerinnen und Schülern gemäß der KMK-Bildungsstandards mit einem kontextorientierten Biologieunterricht zu fördern. Die empirische Überprüfung erfolgte hier in unterschiedlicher Weise (Bayrhuber, Bögeholz, Eggert et al., 2007; Bayrhuber, Bögeholz, Elster et al., 2007). So wählte beispielsweise Grube (2011) Aufgaben im offenen Antwortformat, um Problemlöseprozesse zu erfassen und damit eine Überprüfung der Struktur und Entwicklung zu gewährleisten. Reitschert und Höble (2007) nutzten Interviews, um Ausdifferenzierungen und Abstufungen in den einzelnen Teilkompetenzen zu identifizieren. Neben qualitativen Ansätzen wurden auch quantitative Zugänge gewählt. Schmiemann (2010) setzte zur Überprüfung eines Kompetenzstrukturmodells im Kompetenzbereich *Fachwissen* Mehrfach-Wahlaufgaben ein, die auf der Basis offener Formate entstanden sind. Die entwickelten Aufgaben in den Kompetenztests sind vor ihrem Einsatz zur Überprüfung von Kompetenzmodellen unter anderem daraufhin überprüft worden, ob sie als Indikatoren der spezifischen Kompetenz interpretiert werden können (Schmiemann, 2010).

Zur Überprüfung von Kompetenzstrukturmodellen und Kompetenzniveaumodellen setzten die verschiedenen Forschungsgruppen ihre entwickelten Kompetenztests bei Schülerinnen und Schülern der Sekundarstufe I in Querschnittstudien ein (Eggert, 2008; Schmiemann, 2010). Sollte darüber hinaus auch eine Aussage über Entwicklungsprozesse getroffen werden, sind Längsschnitterhebungen durchgeführt worden (Grube, 2011). Qualitativ erhobene Daten wurden mithilfe von zuvor entwickelten Kodierleitfäden inhaltsanalytisch ausgewertet (Eggert, 2008; Grube, 2011; Kramer, 2009; Reitschert & Höble, 2007). Darüber hinaus fanden bei Studien mit großer Stichprobenanzahl Verfahren der Item-Response-Theorie Verwendung, um unter anderem die Dimensionalität der theoretisch fundierten Kompetenzmodelle zu untersuchen. Als Messmodelle eignen sich bei offenen Formaten vor allem Partial-Credit-Modelle (Eggert & Bögeholz, 2010; Grube, 2011). Das Forschungsdesign der vorliegenden

Arbeit, das ebenfalls die Überprüfung eines Kompetenzmodells zum Ziel hat, orientiert sich an den Vorgehensweisen dieser Studien.

2.2 Perspektiven auf Modelle

So wie das Wort *Kompetenz* ist auch das Wort *Modell* im alltäglichen Leben, im Schulalltag und in der Wissenschaft allgegenwärtig und hat vielfältige Verwendungen. Kinder beispielsweise spielen mit Puppenhäusern und/oder Modelleisenbahnen. Ausgewählte Kundinnen oder Kunden werden bei der Friseurin oder beim Friseur zum Frisurenmodell und zukünftige Eigentümerinnen und Eigentümer sind bei der Planung ihres Eigenheims mit Architekturmodellen konfrontiert.

Die Verwendung von Modellen ist vielfältig und auch im Bereich der Naturwissenschaften gibt es Unterschiede (vgl. Kauertz, Fischer, Mayer, Sumfleth & Walpuski, 2010). So spielen in der Biologie häufig gegenständliche Modelle, beispielsweise 3D-Modelle der DNA oder Ökosystemmodelle eine Rolle, die Funktionen, Strukturen oder dynamische Prozesse repräsentieren. In der Physik und der Chemie sind dagegen vorwiegend abstrakte Modelle, wie Formeln oder Atommodelle, im Einsatz, die zur Theoriebildung herangezogen werden (Beerenwinkel & Parchmann, 2008; Mikelskis-Seifert & Fischler, 2003; Mikelskis-Seifert & Kasper, 2011). Es ist anzunehmen, dass diese unterschiedliche Verwendung daraus resultiert, dass unterschiedliche Sinne bzw. Hilfsmittel herangezogen werden, um die für das Fach relevanten Phänomene wahrzunehmen. Die Biologie zum Beispiel befasst sich hauptsächlich mit konkreten lebenden Phänomenen, wohingegen sich die Physik und die Chemie auf abstrakte Phänomene (z. B. Theorien, Moleküle) beziehen.

Gründe für die Allgegenwart und die vielfältige Verwendung des Modellbegriffs sind möglicherweise auch in der langen Tradition des Wortes und des Konzeptes *Modell* begründet (Mahr, 2003, 2008). Das Wort *Modell* entstammt der Architektur, so ist in Vitruvs³ Modell der antiken Bauweise der *modulus* eine Maßeinheit zur Vermessung von Gebäuden. Später leiteten sich daraus Begriffe wie *Modul* (eine Einheit rechteckiger Gitter für Grundrisse), *Model* (Form, z. B. Backform) und *modello* (Entwurf bzw. verkleinerter dreidimensionaler Bau) ab (Mahr, 2003, 2008). Das Spektrum des Modellbegriffs erstreckt sich dabei „von

³ Vitruvius Pollio, berühmter römischer Architekt des ersten Jahrhunderts nach Christus

einer konkreten Technik hin zu dem methodischen Abstraktum, als das wir es heute verstehen“ (Mahr, 2008, S. 190).

Mit dieser Bandbreite an unterschiedlichen Bedeutungen und Verwendungen ist es schwer, gemeinsame Merkmale zu formulieren, die einen Gegenstand zweifelsfrei als Modell erkennbar machen (Mahr, 2008). Eine Gemeinsamkeit weisen die als Modelle bezeichneten Objekte allerdings auf: Sie sind von einem Subjekt zu einem Modell erklärt worden und transportieren eine Idee (Mahr, 2008; Mittelstraß, 2004; Stachowiak, 1973). Somit ist das Subjekt für die Auffassung eines Gegenstands als Modell verantwortlich. Dies bedeutet, dass jedes – auch alltägliche – Objekt von einem Subjekt zu einem Modell erklärt werden kann: Eine Person kann beispielsweise das Verkehrsnetz einer Stadt als Modell heranziehen, um damit ihre Idee von der Struktur des Gehirns zu verdeutlichen. Diese Idee beinhaltet die Vorstellung, dass das Nervennetz komplex ist und eine Vielzahl von Verzweigungen und Verbindungen aufweist. Aus wissenschaftstheoretischer Sicht lassen sich hinsichtlich des Transports dieser Idee zwei Perspektiven auf Modelle unterscheiden: eine mediale und eine methodische (Gilbert, 1991; Mahr, 2008; Oh & Oh, 2011; Terzer, 2013). Diese beiden Perspektiven sind grundlegend für die Entwicklung der Niveaus des Kompetenzmodells der Modellkompetenz (Upmeier zu Belzen & Krüger, 2010), weshalb sie näher beschrieben werden (vgl. S. 24ff. dieser Arbeit).

Die Perspektive von Lernenden auf Modelle ist eine weitere Grundlage des Kompetenzmodells von Upmeier zu Belzen und Krüger (2010). Untersuchungen zeigen, dass Schülerinnen und Schüler Modelle eher aus einer medialen und weniger aus einer methodischen Perspektive betrachten (z. B. Grosslight et al., 1991; Harrison & Treagust, 2000; Prenzel et al., 2004; vgl. S. 27ff. dieser Arbeit). Sie sehen Modelle als Mittel an, um bekannte Sachverhalte zu veranschaulichen sowie zugänglich und vermittelbar zu machen (Ingham & Gilbert, 1991). Bestandteil einer umfassenden Modellkompetenz ist unter anderem eine mediale und methodische Perspektive auf Modelle. Somit sind Schülerinnen und Schüler bei der Entwicklung einer umfassenden Modellkompetenz zu unterstützen (vgl. Fleige, Seegers, Upmeier zu Belzen & Krüger, 2012a; Fleige et al., 2012b; Füssenich et al., 2013; Grünkorn, Lotz et al., 2014). Im Hinblick auf die Förderung der Modellkompetenz liefern Kompetenzmodelle Lehrkräften eine Struktur der zu fördernden Aspekte (vgl. Klieme, Avenarius et

al., 2007). Die Literatur bietet hierzu unterschiedliche Strukturierungen an (vgl. S. 34ff. dieser Arbeit).

2.2.1 Wissenschaftstheoretische Sichtweisen von Modellen

Die Beobachtung eines Phänomens kann dazu führen, dass wissenschaftliche Fragen aufgeworfen werden (Frigg & Hartmann, 2012). Singer und Nicolson beispielsweise (1972) stellten sich nach der Beobachtung eines Versuchs zur Biomembran, in dem sich markierte Proteine als beweglich erwiesen, folgende Frage: Wie muss eine Biomembran geschaffen sein, damit eine solche Fluidität gegeben ist?

Ausgehend von dieser Wahrnehmung wird zunächst ein internes gedankliches Modell (mentales Modell) konstruiert (vgl. Boulter & Buckley, 2000; Johnson-Laird, 1983; Vosniadou, 2002). In der Biologiedidaktik wird dieses mentale **Modell als „Original“** (Kattmann, 2006, S. 330) bezeichnet und in der Wissenschaftstheorie wird es oft **„target“** oder **„referent“** (Oh & Oh, 2011, S. 1113) genannt (vgl. Terzer, 2013). Am Beispiel Biomembran entwickelten Singer und Nicolson (1972) ein mentales Modell einer Biomembran als Flüssig-Mosaik-Modell (*fluid mosaic model*). Bei diesem Modell bewegen sich mosaikartig angeordnete Proteine auf und in einer flüssigen, in Bewegung befindlichen Phosphorlipidschicht. Dieses mentale Modell repräsentiert somit das Ergebnis ihrer Dateninterpretation bzw. zeigt, wie sich Singer und Nicolson (1972) die Biomembran vorstellten. Im Sinne einer konstruktivistischen Sichtweise bedeutet dies, dass sich mentale Modelle von Subjekt zu Subjekt unterscheiden können und für das Subjekt lediglich empirisch adäquat sein müssen (*constructive empiricism*, van Fraassen, 1980).

Weiterhin kann ein bereits existierendes mentales Modell durch zusätzliche Beobachtungen und Informationen verändert werden, und daraus kann sich ein neues mentales Modell entwickeln (Vosniadou, 2002). Vosniadou (2002, S. 4) schreibt: „[...] it is assumed that most mental models are constructed at the spot to deal with the demands of specific situations, although it is possible that some mental models may be stored in long-term memory“. Mentale Modelle sind somit einem dynamischen Prozess unterworfen und veränderbar, weshalb anzunehmen ist, dass sie auf unterschiedlich stabilen kognitiven Strukturen basieren (Greca & Moreira, 2000; Vosniadou, 2002). Die Veränderbarkeit von mentalen Modellen lässt sich auch am bereits herangezogenen Bei-

spiel der Biomembran zeigen. So entwickelten Kusumi und Sako (1996) aus dem Flüssig-Mosaik-Modell ihr mentales Modell der abgegrenzten Bezirke (*fence-and-picket-model*) und Simons und Ikonen (1997) parallel dazu ihr Lipid-Flöß-Modell (*lipid-raft-model*). Dem ersten Modell liegt die Hypothese zugrunde, dass das Cytoskelett **bestimmte Bereiche mit Proteinen** „einzäunen“ kann, wodurch die Beweglichkeit der Proteine eingeschränkt wird. Einige Proteine übernehmen die Funktion von „**Zaunpfosten**“, andere wiederum können diese abgegrenzten Bezirke überspringen (Abbot, 2005; Kusumi & Sako, 1996; Suzuki, Ritchie, Kajikawa, Fujiwara & Kusumi, 2005). Simons und Ikonen (1997) postulieren hingegen, es gebe in den Biomembranen dynamische Strukturen (sogenannte Flöße, engl. rafts). Diese seien beweglich, veränderten sich hinsichtlich der Zusammensetzung und könnten einzelne Proteine aufnehmen oder abgeben. Diese alternativen mentalen Modelle der Biomembran sind allerdings zunächst nur gedanklich konstruiert und anderen Forschenden nicht zugänglich.

Um sich über die Ideen zu einem wahrgenommenen Phänomen, wie hier zur Biomembran, mit anderen Forschenden auszutauschen, ist es hilfreich, diese zu visualisieren (Norman, 1983; Oh & Oh, 2011). Erst mit einer Visualisierung ist es möglich, ein mentales Modell (im Weiteren: Original) kommunizierbar zu machen (Nersessian, 1992). Das Original bildet die Basis einer solchen Visualisierung bzw. Konstruktion eines gegenständlichen, konzeptuellen Modells (Abb. 3). Bei dieser Konstruktion übernehmen die Modellierenden eine entscheidende Rolle. Sie bestimmen, welche Merkmale des Originals für die Repräsentation ihrer Idee relevant sind. Aus diesem Grund ist ein Modell immer subjekt- und zweckgebunden (Bailer-Jones, 2002; Giere, 2001, 2010; Halloun, 2001; Mahr, 2008; Odenbaugh, 2005; Stachowiak, 1973). Bei der Herstellung eines konzeptuellen Modells werden einerseits bestimmte Merkmale des Originals vereinfacht, und damit wird die Aufmerksamkeit auf relevante Merkmale gelenkt (Nersessian, 1992). Stachowiak (1973) bezeichnet dies als „**Verkürzungsmerkmal**“ (S. 131) von Modellen. Andererseits repräsentiert das konzeptuelle Modell ausgewählte Merkmale des Originals (Giere, 2010; „**Abbildungsmerkmal**“, Stachowiak, 1973, S. 131). Dabei ist das konzeptuelle Modell hinsichtlich der Repräsentation dieser Merkmale möglichst genau und stimmt mit den zu diesem Zeitpunkt vorliegenden wissenschaftlichen Erkenntnissen überein (Greca & Moreira, 2000; Norman, 1983). Es ist **festzuhalten**: “[...] if a

model were exactly like its target, it would not be a model **but a copy**" (van der Valk, van Driel & de Vos, 2007, S. 471). Vor diesem Hintergrund kann das konzeptuelle Modell als Modell **von** etwas betrachtet werden, weil es ausgewählte Ausschnitte des Originals repräsentiert (Mahr, 2003, 2008).

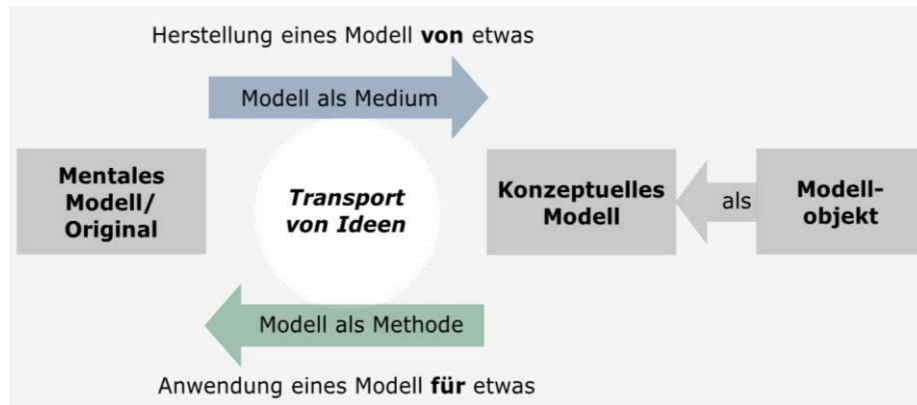


Abb. 3: Wissenschaftstheoretische Perspektiven auf Modelle (verändert nach Terzer, 2013)

Mit der Herstellung eines konzeptuellen Modells als Modell **von** etwas ist die mediale Perspektive auf Modelle zwar grundlegend, aber noch nicht hinreichend beschrieben (Gilbert, 1991; Mahr, 2008, 2009; Oh & Oh, 2011). Das liegt daran, dass die Modellierenden bei der Konstruktion auch eine Entscheidung darüber treffen, welche Darstellungsform das konzeptuelle Modell besitzen soll (Mahr, 2004, 2008). Das bedeutet, dass entschieden werden muss, welche Erscheinungsform (z. B. Text, Grafik, Körper) gewählt wird, welche Farbe und Dimension das konzeptuelle Modell haben und aus welchem Material es bestehen soll. Singer und Nicolson (1972) entschieden sich bei der Veröffentlichung ihres Flüssig-Mosaik-Modells beispielsweise für eine Schwarz-Weiß-Zeichnung. Diese ergänzende Perspektive beschreibt die Betrachtung des „**Gegenstand[s]**“ (Mahr, 2004, S. 11) bzw. des Modellobjekts (Mahr, 2004, 2008). Somit wird die mediale Perspektive um die Betrachtung des Modellobjekts erweitert (Abb. 3).

Neben dieser Visualisierung ihrer Idee vom Aufbau der Biomembran nutzten Singer und Nicolson (1972) ihr konzeptuelles Modell auch dazu, Hypothesen über die Permeabilität der Biomembran zu formulieren und diese zu überprüfen. Auf diese Weise können Modelle dazu herangezogen werden, Hypothesen über das Original zu testen und mit dem Modell Erkenntnisse über das Original zu gewinnen. Ferner lassen sich die Gültigkeit von Theorien überprüfen und

neue Ideen oder Theorien über das Original entwickeln (Abel, 2008; Giere, 1988; Mahr, 2003, 2004; Morrison & Morgan, 1999; Stachowiak, 1973; van Fraassen, 1980). Dabei kann sich die Vorstellung von den modellierten Merkmalen des Originals verändern (Grosslight et al., 1991; Mahr, 2003, 2004; Oh & Oh, 2011; Stachowiak, 1973).

Bei einer solchen Anwendung des Modells analysieren und beurteilen die Modellierenden die zugrunde liegende Idee im konzeptuellen Modell. Hierbei fragen die Forschenden nicht danach, ob die Modelle richtig oder falsch sind, sie hinterfragen aber, ob mit dem Modell das wahrgenommene Phänomen erklärt oder das Verhalten des Originals vorhergesagt werden kann (Crawford & Cullin, 2004). Demzufolge prüfen Wissenschaftlerinnen und Wissenschaftler bei der Anwendung die Passung zwischen dem Modell und dem Phänomen (Giere, 2001, 2004). Dadurch, dass ein Phänomen häufig komplex ist und zudem oft nur mit dem Einsatz von Hilfsmitteln (z. B. Mikroskop, Röntgengerät, chemischen Untersuchungen) wahrgenommen werden kann, kann eine Passung mit dem Modell nicht immer abschließend beurteilt werden (Giere, Bickle & Mauldin, 2006; van der Valk et al., 2007). Infolgedessen sind Modelle Mediatoren, weil sie zwischen einer Idee oder einer Theorie und dem jeweiligen Phänomen vermitteln (Koponen, 2007; Morrison & Morgan, 1999; Rotbain, Marbach-Ad & Stavy, 2006). Dadurch ermöglichen sie es, neues Wissen über die Welt zu generieren (Giere et al., 2006). Im Hinblick auf die Anwendung von Modellen als Methode sind Modelle somit auch als Modelle *für* etwas zu verstehen (Gilbert, 1991; Mahr, 2008, 2009; Oh & Oh, 2011; Abb. 3).

Zusammengefasst sind Modelle aus wissenschaftstheoretischer Sicht **Medien** zur Veranschaulichung und **Methoden** zur Erkenntnisgewinnung (Gilbert, 1991; Mahr, 2008, 2009; Oh & Oh, 2011; Abb. 3). **In ihrer Funktion als „a bridge or mediator connecting a theory and a phenomenon“ (Oh & Oh, 2011, S. 1114)** unterstützen sie den Prozess, Ideen oder Theorien über die Welt zu repräsentieren und zu entwickeln (Giere et al., 2006).

2.2.2 Empirische Perspektiven auf Modelle

In einigen empirischen Studien zu Modellen und zur Modellbildung wurden die Vorstellungen von Schülerinnen und Schülern (z. B. Grosslight et al., 1991; Krell, Upmeier zu Belzen et al., 2012b; Schwarz et al., 2009; Schwarz & White, 2005; Treagust, Chittleborough & Mamiala, 2002; Terzer, 2013; Trier &

Upmeier zu Belzen, 2009) sowie von angehenden wie erfahrenen Lehrkräften (z. B. Crawford & Cullin, 2004, 2005; Justi & Gilbert, 2002, 2003; van Driel & Verloop, 1999) untersucht. Unter dem Begriff *Vorstellungen* sind hier „subjektive gedankliche Prozesse“ (Gropengießer, 2003, S. 13) zu verstehen, die „individuelle Leistungen“ (Weitzel & Gropenießer, 2009, S. 288) beschreiben. Diese Prozesse sind komplex und basieren auf kognitiven Strukturen, die entweder bereits vorliegen oder in bestimmten Situationen konstruiert werden (Greca & Moreira, 2000; Vosniadou, 2002). Damit sind Vorstellungen an Kontexte gebunden und veränderbar (Sins, Savelsbergh, van Joolingen & van Hout-Wolters, 2009). Darüber hinaus entwickeln sich Vorstellungen unterschiedlich, da die ihnen zugrunde liegenden kognitiven Strukturen unterschiedlich stabil sind (*construction-in-interaction framework*, Boersma & Geraedts, 2012). Nach dieser Beschreibung entsprechen Vorstellungen dem Verständnis von kognitiven Kompetenzen (vgl. S. 15ff. dieser Arbeit) und dem von mentalen Modellen (vgl. S. 24ff. dieser Arbeit).

Instrumente zur Erfassung der Vorstellungen von Modellen

Um Vorstellungen von Modellen und der Modellbildung zu erfassen, sind unterschiedlichste Herangehensweisen und Instrumente entwickelt worden. Grosslight et al. (1991) beispielsweise führten Interviews mit Schülerinnen und Schülern der siebten und der elften Jahrgangsstufe sowie mit Expertinnen und Experten durch. In diesen Interviews wurden den Personen spezifische Objekte (z. B. Spielflugzeug, U-Bahn-Plan) gezeigt, um festzustellen, ob sie diese als Modelle ansehen. Ansonsten bezogen sich die in den Interviews gestellten Fragen nicht auf spezifische inhaltliche Situationen, sie waren also kontextunabhängig. Eine ähnliche Vorgehensweise wählten Justi und Gilbert (2003) sowie Trier und Upmeier zu Belzen (2009). Erstere erfassten kontextunabhängig mithilfe eines halbstrukturierten Interviewleitfadens die Vorstellungen von Lehrkräften und Letztere die Vorstellungen von Schülerinnen und Schülern.

Neben Interviews wurden auch andere offene Ansätze gewählt, um die verschiedenen Vorstellungen von Schülerinnen und Schülern, die diese von Modellen und von dem Modellbildungsprozess haben, zu identifizieren. Sins, Savelsbergh, van Joolingen und van Hout-Wolters (2009) setzten bei Lernenden der elften Jahrgangsstufe Aufgaben mit einem offenen Antwortformat ein. Die Kon-

texte in den Aufgaben orientierten sich an denen der Modellierungsaufgabe, mit der die Lernenden zuvor gearbeitet hatten.

Treagust et al. (2002) entwickelten dagegen für ihre Studie mit Schülerinnen und Schülern ein quantitatives Instrument namens *Students' Understanding of Models in Science* (SUMS). Dieses Instrument basiert auf den Daten einer Vorgängerstudie (Treagust, Chittleborough & Mamiala, 2001) sowie den Daten der Studie von Grosslight et al. (1991). Der SUMS-Fragebogen besteht aus kontextunabhängig formulierten Items mit einer fünfstufigen Likert-Skala.

Um die Validität der Ergebnisse zu erhöhen, entschieden sich einige Forschende dafür, qualitative und quantitative Ansätze zu kombinieren. So nutzten Treagust, Chittleborough und Mamiala (2004) in ihrer Studie unter anderem Interviews sowie quantitative Instrumente, wie den VOMMS-Fragebogen (*My Views of Models and Modelling in Science*) und das Instrument *Molecular Representations*. Letzteres enthält Items mit einer fünfstufigen Likert-Skala, die sich auf den Zweck von Modellen beziehen. Der VOMMS-Fragebogen ist kontextunabhängig formuliert und fordert die Lernenden dazu auf, sich zwischen zwei Aussagen zu entscheiden (FC-Items). Auch Schwarz und White (2005) kombinierten in ihrer Interventionsstudie zu Modellen kontextunabhängige geschlossene Formate (MC-Items, Kategorisierungsaufgaben und Richtig-/Falsch-Aufgaben) mit Interviews. Im Gegensatz zu den geschlossenen Formaten bezogen sich die Interviews auf einen Kontext. Van Driel und Verloop (1999) setzten neben kontextunabhängigen Aufgaben im offenen Format auch Items mit einer vierstufigen Likert-Skala ein, um die Vorstellungen von Lehrkräften zu erheben. Andere Forschende, wie Crawford und Cullin (2005), kombinierten verschiedene offene Formate. Sie nutzten beispielsweise in ihrer Interventionsstudie mit angehenden Lehrkräften einen offenen Fragebogen sowie Interviews, um die Vorstellungen vor und nach der Intervention zu erfassen. Während die Fragen im Pretest kontextunabhängig waren, bezogen sich die Fragen nach der Intervention auf den konkreten Inhalt der Intervention (Thema **See**). Es hat sich also gezeigt, dass Vorstellungen von Modellen und von der Modellbildung auf unterschiedliche Weise erfassbar sind. Neben verschiedenen Formaten und deren Kombinationen lassen sich kontextabhängige und kontextunabhängige Ansätze voneinander unterscheiden. Sins et al. (2009) kritisieren allerdings die kontextunabhängigen Ansätze. Sie begründen dies mit der Kontextgebundenheit epistemologischer Vorstellungen. Krell, Upmeier zu Belzen et

al. (2012b, 2014c) konstatieren in ihrer Studie mit kontextabhängigen und kontextunabhängigen FC-Items, dass einerseits Unterschiede zwischen diesen beiden Ansätze festzustellen sind, dass andererseits die Ergebnisse aber zeigen, dass sich die kontextgebundenen FC-Items bezüglich der Performanz von Lernenden unterscheiden und demzufolge Schwierigkeiten hervorrufen. Diese Befunde sind bei der Operationalisierung der Teilkompetenzen zu berücksichtigen.

Vorstellungen von Schülerinnen und Schülern von Modellen

Unter Verwendung dieser verschiedenen Instrumente wurden Vorstellungen von Modellen und der Modellbildung erfasst. Bei Schülerinnen und Schülern zeigte sich in Bezug auf die wissenschaftstheoretischen Sichtweisen eine hauptsächlich mediale Perspektive auf Modelle (z. B. Grosslight et al., 1991; Krell, Upmeier zu Belzen et al. 2012b; Schwarz et al., 2009; Schwarz & White, 2005; Treagust et al., 2002; Terzer, 2013; Trier & Upmeier zu Belzen, 2009). Diese äußerte sich hinsichtlich der *Beziehung von Modell und Original (mentales Modell)* darin, dass viele Schülerinnen und Schüler in Bezug auf Modelle an konkrete Objekte bzw. Replikat, wie Spielzeugautos oder Architekturmodelle, denken. Nur wenige Schülerinnen und Schüler beziehen sich auf abstrakte Objekte (mathematische Modelle, Formeln), in denen Modelle als Repräsentationen von Ideen verstanden werden (*kinds of models – examples of models*, Grosslight et al., 1991). Betrachten und beschreiben Lernende die Beziehung von Modell und Original, betonen sie, dass Modelle in Farbe, Dimension und Material dem Original ähnlich sein sollen (*kinds of models – relationship, designing and creating models – what the real thing is*, Grosslight et al., 1991). Nur wenige Lernende erwähnen die Rolle der Modellierenden in diesem Prozess, die aber maßgeblich den Zweck des Modells bestimmen (*kinds of models – awareness of modeler*, Grosslight et al., 1991). Neben Grosslight et al. (1991) kommen auch andere Studien zu ähnlichen Ergebnissen. Sie stellen fest, dass Lernende Modelle häufig als zwei- oder dreidimensionale Duplikate der Welt wahrnehmen (*definition of a model*, Schwarz & White, 2005; *scientific models as exact replicas*, Treagust et al., 2002; *Eigenschaften*, Trier & Upmeier zu Belzen, 2009). Diese Vorstellung ist mit einem naiv-realistischen Wissenschaftsverständnis zu erklären (Hofer & Pintrich, 1997).

Hinsichtlich des Vorhandenseins *verschiedener Modelle eines Originals* führen Grosslight et al. (1991, *multiple models*) und Treagust et al. (2002, *scientific models as multiple representations*) an, dass ein Großteil der Schülerinnen und Schüler dieser Möglichkeit zustimmt. Lernende begründen alternative Modelle eines Originals damit, dass sich das Original mithilfe verschiedener Modelle aus unterschiedlichen Blickwinkeln (innen/außen) betrachten lässt (*designing and creating models – depends on the view*, Grosslight et al., 1991). Eine weitere verbreitete Begründung ist, dass damit auf verschiedene inhaltliche Schwerpunkte (Struktur/Funktion) fokussiert werden kann (*designing and creating models – depends on what is major/minor, basics/details, little things/big things, main things*, Grosslight et al., 1991). In diesem Zusammenhang heben Schülerinnen und Schüler hervor, dass sich mithilfe alternativer Modelle ein besseres Verständnis des Originals erreichen lässt, weil das Original übersichtlicher dargestellt werden kann und an Klarheit gewinnt (*multiple models – yes, to show different views of the same entity, purpose of models – accessibility or clarity*, Grosslight et al., 1991; *scientific models as multiple representations*, Treagust et al., 2002). Zudem erwähnen Lernende, „dass mit alternativen Modellen verschiedene Lerntypen angesprochen werden können“ (*alternative Modelle*, Trier & Upmeier zu Belzen, 2009, S. 32). Diese Vorstellung äußern Schülerinnen und Schüler auch in der Studie von Treagust et al. (2002). Die methodische Perspektive auf Modelle, in der Modelle verschiedene Hypothesen über das Original repräsentieren und die Hypothesen damit getestet werden können, beschreiben Schülerinnen und Schüler selten (*multiple models – two or more different ideas or ways of explaining „it“, yes, you can make different models to test an entity*, Grosslight et al., 1991; *alternative Modelle*, Trier & Upmeier zu Belzen, 2009).

Geht es um den *Zweck von Modellen*, dann zeigen Schülerinnen und Schüler vielfältige Vorstellungen (vgl. Krell, Upmeier zu Belzen et al., 2012b). So unterscheiden sie zwischen in der Schule und in der Wissenschaft eingesetzten Modellen (*Zweck von Modellen*, Trier & Upmeier zu Belzen, 2009). Während im schulischen Kontext mit Modellen bekannte Zusammenhänge demonstriert werden, dienen sie im wissenschaftlichen Kontext dazu, unbekannte Sachverhalte zu entdecken und zu begreifen (*Zweck von Modellen*, Trier & Upmeier zu Belzen, 2009). Ausgehend davon ist es eine verbreitete Vorstellung von Lernenden, dass Modelle Merkmale des Originals oder Zusammenhänge bzw. Pro-

zesse veranschaulichen (*purpose of models – communication*, Grosslight et al., 1991; *Zweck von Modellen*, Trier & Upmeier zu Belzen, 2009). Weiterhin spielen bei Schülerinnen und Schülern Modelle als Mittel der Kommunikation und der Verständlichmachung von Sachverhalten eine große Rolle. In diesem Sinne verbessern Modelle das Verständnis des Originals und erleichtern die Vermittlung von Wissen (*purpose of models – communication, learning, understanding, designing and creating models – communication*, Grosslight et al., 1991; *Zweck von Modellen*, Trier & Upmeier zu Belzen, 2009). Kattmann (2006) umschreibt dies als „denkökonomische Funktion“ (S. 332) von Modellen. Darüber hinaus dienen Modelle dazu, Phänomene zugänglich bzw. direkt erfahrbar zu machen (*purpose of models – accessibility or clarity*, Grosslight et al., 1991). Auf diese Weise fungieren Modelle als Mittel der Zugänglichmachung von Phänomenen in der Welt. Ingham und Gilbert (1991) beschreiben diese didaktisch orientierte Funktion von Modellen in ihrer Untersuchung mit Studierenden der Chemie als „the use of models for personal purposes“ (S. 197) und als „the use of models for the purposes of others“ (S. 198).

Neben diesen Vorstellungen nennen Schülerinnen und Schüler auch den alltäglichen Nutzen von Modellen. Hier dienen Modelle als Referenz (z. B. ein Bauplan), als Spielzeug, als Vorbild für Personen oder Dekorationen (*purpose of models – reference or example*, Grosslight et al., 1991). Diese vielfältigen Vorstellungen beschreiben alltägliche Vorstellungen und zeigen eine mediale Sicht von Modellen. Nur vereinzelt wird das Modell als ein Mittel der Erkenntnisgewinnung herangezogen, mit dem sich Ideen und Hypothesen untersuchen lassen, um neue Erkenntnisse hinsichtlich des Originals zu gewinnen und Voraussagen über das Original zu tätigen (*purpose of models – testing*, Grosslight et al., 1991; *Zweck von Modellen*, Trier & Upmeier zu Belzen, 2009).

Den Prozess des *Testens und Änderns von Modellen* verstehen Schülerinnen und Schüler als unsystematisch und zufällig (Grosslight et al., 1991). Dies äußert sich darin, dass sie Modelle allgemein und ungerichtet testen, indem sie etwas ausprobieren oder schauen, was passiert (*purpose of models – testing*, Grosslight et al., 1991). Manche Schülerinnen und Schüler beschreiben, dass die Ähnlichkeit zwischen Modell und Original geprüft werden muss (*designing and creating models – should be exact, smaller, propotional*, Grosslight et al., 1991; *Testen von Modellen*, Trier & Upmeier zu Belzen, 2009). Ein weiteres Kriterium, welches zur Überprüfung eines Modells herangezogen wird, ist die

Verständlichkeit (Schwarz & White, 2005). Nur selten beschreiben Lernende die Vorstellung, dass sich mit Modellen Ideen oder Hypothesen überprüfen lassen (*purpose of models – testing*, Grosslight et al., 1991; *purpose of models*, Treagust et al. 2002). Diese Befunde weisen darauf hin, dass Lernende das Testen von Modellen nicht als einen entscheidenden Faktor für die Entwicklung und Revision von Ideen und Theorien in Betracht ziehen und erkennen (Treagust et al., 2002).

Als Folge einer Testung kann es zu Veränderungen am Modell kommen. Nur einige wenige Schülerinnen und Schüler äußern in diesem Zusammenhang die Vorstellung, Modelle seien unveränderlich (*changing a model – other or „no“*, Grosslight et al., 1991). Ein Großteil der Schülerinnen und Schüler meint, dass Forschende Modelle verändern (Grosslight et al. 1991; Treagust et al., 2002). Die Erklärungen von Veränderungsprozessen sind allerdings bei jüngeren Schülerinnen und Schüler sehr vage. So begründen sie Veränderungen am Modell lediglich damit, dass Fehler im Modell identifiziert wurden (*changing a model – something wrong with model or how model was made*, Grosslight et al., 1991). Ältere Schülerinnen und Schüler begründen dies dagegen vielfältiger und genauer. Vielfach wird das Verändern von Modellen mit einer mangelnden strukturellen und/oder funktionalen Passung zwischen Modell und Original oder mit neuen Erkenntnissen hinsichtlich des Originals begründet (*changing a model – model not right because new information is found, yes, model changes in accordance with how purpose or need was matched*, Grosslight et al., 1991; *Ändern von Modellen*, Trier & Upmeier zu Belzen, 2009). Andere würden ein Modell verändern, damit es schöner wird (*changing a model – yes, for aesthetic reasons*, Grosslight et al., 1991). Nur wenige bis keine Schülerinnen und Schüler beziehen sich in ihrer Begründung auf die Falsifikation von Hypothesen. Dies bedeutet, dass sich bei der Anwendung eines Modells eine Hypothese über das Original als nicht haltbar erwiesen hat (Grosslight et al., 1991; Trier & Upmeier zu Belzen, 2009).

Insgesamt ist festzustellen, dass Schülerinnen und Schüler im Gegensatz zu Wissenschaftlerinnen und Wissenschaftlern Modelle vorwiegend als Mittel der Veranschaulichung und weniger als Mittel der Erkenntnisgewinnung verstehen und nutzen (z. B. Grosslight et al., 1991; Krell, Upmeier zu Belzen et al., 2012b; Schwarz et al., 2009; Schwarz & White, 2005; Treagust et al., 2002; Terzer, 2013; Trier & Upmeier zu Belzen, 2009). Diese unterschiedlichen Vor-

stellungen gehen oft mit einer didaktisch orientierten Funktion von Modellen einher. Lernende können mit Modellen nicht greifbare oder nicht sichtbare Phänomene zugänglich machen, um diese besser zu verstehen (Grosslight et al., 1991; Ingham & Gilbert, 1991). Die Befunde deuten bislang darauf hin, dass Lernende die didaktisch orientierten Perspektiven mit konkreten Vorstellungen und einer naiv-realistischen Sichtweise auf die Welt verbinden (Grosslight et al., 1991). Das bedeutet, dass mit einem Modell beispielsweise ein Phänomen in Form eines konkreten 3-D-Modells visualisiert wird. Auf diese Weise ist eine Person in der Lage, das Modell in die Hand zu nehmen und etwas über das Phänomen zu lernen (Ingham & Gilbert, 1991). Damit überwiegt die **von Kattmann (2006) beschriebene „Anschauungsfunktion“ (S. 332)** von Modellen, in der Modelle der Veranschaulichung von Strukturen und Prozessen dienen. In diesen didaktischen Sichtweisen liegt allerdings auch das Potenzial dazu, Modelle als Mediatoren wahrzunehmen, die zwischen Theorie und Phänomen vermitteln (Oh & Oh, 2011). **Damit wird ihre „heuristische Funktion“ (Kattmann, 2006, S. 332)** in den Blick genommen, in der Modelle durch ihren hypothetischen Charakter Phänomene erklären können. Dies verlangt ein eher abstraktes Verständnis von Modellen und eine konstruktivistische Sichtweise (van Fraassen, 1980). Daher ist anzunehmen, dass diese didaktisch orientierten Perspektiven, aber auch weitere Besonderheiten bei den Antworten der Lernenden für die Entwicklung von Modellkompetenz relevant sind. Sie werden deshalb in der vorliegenden Arbeit in den Blick genommen (vgl. S. 104ff. dieser Arbeit).

2.2.3 Strukturierungen der Vorstellungen von Modellen

Über die Möglichkeit hinaus, Besonderheiten bei den Antworten der Lernenden zu identifizieren, können Strukturierungen der Vorstellungen von Modellen dabei helfen, Fördermaßnahmen für Schülerinnen und Schüler zu entwickeln. Solche Ansätze strukturieren theoretisch und/oder empirisch abgeleitete Vorstellungen in verschiedene Aspekte und/oder Ausprägungen. In der Literatur zu Modellen und zur Modellbildung sind bereits einige mögliche Strukturierungen beschrieben (z. B. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Leisner-Bodenthin, 2006; Louca, Zacharia, Michael & Constantinou, 2011; Meisert, 2008; Oh & Oh, 2011; Schwarz et al., 2009; Treagust et al., 2002).

Grosslight et al. (1991) identifizierten in ihrer Interviewstudie mit 33 Lernenden der siebten Jahrgangsstufe, 22 Lernenden der elften Jahrgangsstufe sowie mit vier Expertinnen und Experten fünf Aspekte von Modellen: *kinds of models*, *multiple models for the same thing*, *purpose of models*, *designing and creating models* und *changing a model*. Zudem generierten sie aus den Interviewanalysen drei generelle Niveaus, die über alle Aspekte hinweg gelten. Diese drei Niveaus sind eng mit verschiedenen Ausprägungen des Wissenschaftsverständnisses verbunden (Carey, Evans, Honda, Jay & Unger, 1989; Hofer & Pintrich, 1997). Dabei reicht das Spektrum von einem naiv-realistischen bis hin zu einem konstruktivistischen Wissenschaftsverständnis: Schülerinnen und Schüler auf **Niveau I** verstehen Modelle als einfache Kopien der Realität, die dem Original in Farbe, Form oder Dimension gleichen sollen. Mit dieser Sicht auf Modelle ist ein naiv-realistisches Wissenschaftsverständnis verbunden. Auf **Niveau II** machen sich die Lernenden bewusst, dass Modelle zweckgebunden konstruiert werden und dass demzufolge die Modellierenden eine wichtige Rolle im Modellierungsprozess spielen. Modelle werden dabei nicht als exakte Duplikate der Originale verstanden, sondern als idealisierte Repräsentationen. Diese Sicht auf Modelle beschreibt ein relativistisches Wissenschaftsverständnis. Lernende auf **Niveau III** haben eine konstruktivistische Sicht auf Modelle. Ihnen ist bewusst, dass das Original erst vermittels der Anwendung des Modells erklärt werden kann. Modelle werden hier als Modelle *für* etwas und somit als **Methode** verstanden, mit deren Hilfe Ideen getestet und Erkenntnisse gewonnen werden können (Grosslight et al., 1991).

Diese drei auf erkenntnistheoretischer Grundlage allgemein formulierten Niveaus von Grosslight et al. (1991) bildeten die Basis für die Interviewstudie von Justi und Gilbert (2003) mit 39 Lehrkräften. Die Studie hatte das Ziel, Vorstellungen von Lehrkräften sowie Muster innerhalb dieser Vorstellungen zu identifizieren. Aus der Analyse der Interviews generierten Justi und Gilbert (2003) zunächst sieben Aspekte von Modellen: *nature*, *use*, *entities*, *uniqueness*, *time*, *predictions* und *accreditation*. Im Gegensatz zur Strukturierung von Grosslight et al. (1991) teilten sie den Aspekt *kinds of models* in die Aspekte *nature* und *entities* auf und den Aspekt *purpose of models* in die Aspekte *use*, *prediction* und *accreditation*. Zudem beschrieben Justi und Gilbert (2003) innerhalb des Aspekts *uniqueness* erstmals explizit die Vorstellung, dass es zu einem Original nur ein Modell geben könne (S. 1375). Mit Blick auf Schülerin-

nen und Schüler sollte empirisch untersucht werden, ob sich diese Vorstellung auch bei dieser Zielgruppe zeigen würden (vgl. S. 100ff. dieser Arbeit). Die Analysen zu Antwortmustern ergaben, dass über die sieben Aspekte hinweg keine Profile identifiziert werden konnten, dass es jedoch innerhalb der jeweiligen Aspekte Auffälligkeiten gab (Justi & Gilbert, 2003, S. 1382). So äußerte ein und dieselbe Person innerhalb des Aspekts *nature* unterschiedliche Sichtweisen und sprach von einem Modell als einer identischen Reproduktion und sogar als einer theoretischen Rekonstruktion (Justi & Gilbert, 2003, S. 1378). Begründungen für diese kontroversen Sichtweisen einer Person lassen sich vor allem im kontextunabhängigen Ansatz dieser Studie finden (vgl. Krell, Upmeyer zu Belzen et al., 2012b, 2014c). Es ist anzunehmen, dass die Idee von unterschiedlichen Modellen bei den Interviewten dazu führte, dass sie damit unterschiedliche Sichtweisen von Modellen verbanden. Auf der Grundlage dieser Analysen konnten Justi und Gilbert (2003) auch keine unterstützenden Befunde für die drei übergreifende Niveaus von Grosslight et al. (1991) feststellen. Im Gegensatz dazu schlugen sie vor, für jeden Aspekt ein entsprechendes Niveau zu definieren (Justi & Gilbert, 2003).

Eine weitere Strukturierung entwickelten Crawford und Cullin (2005) aus einer Analyse der relevanten Literatur zu Modellen (de Jong & van Driel, 2001; Grosslight et al., 1991; Harrison, 2001; Justi & Gilbert, 2001; Smit & Finegold, 1995; van Driel & Verloop, 1999). Sie unterschieden fünf Aspekte von Modellen: *multiple models for the same thing*, *purpose of models*, *designing and creating models*, *validating/testing models* und *changing a model*. Für diese fünf Aspekte leiteten sie aus der Literatur jeweils drei bis vier Niveaus (*limited*, *pre-scientific*, *emerging scientific* und *scientific*) ab. So definierten sie für die Aspekte *purpose of models*, *changing a model* und *multiple models for the same thing* vier Niveaus und für die Aspekte *designing and creating models* und *validating/testing models* drei Niveaus. Dabei beschreiben die Niveaus einen steigenden wissenschaftlichen Reflexionsgrad. Im Gegensatz zu den Strukturierungen von Grosslight et al. (1991) sowie Justi und Gilbert (2003) legten sie keinen separaten Aspekt fest, der sich mit der Beziehung von Modell und Original auseinandersetzt. Außerdem ist diese theoretisch abgeleitete Strukturierung bisher nicht empirisch überprüft worden, weshalb nicht abschließend beurteilt werden kann, ob die definierten Aspekte die Perspektiven von Schülerinnen und Schülern hinreichend berücksichtigen.

So wie Crawford und Cullin (2005) entwickelten auch Oh und Oh (2011) einen theoretischen Strukturierungsansatz aus Analysen verschiedener Publikationen zu Modellen. Sie identifizierten fünf Aspekte: *meanings of a model*, *purposes of modelling*, *multiplicity of scientific models*, *change in scientific models* und *uses of models in the science classroom*. Im Gegensatz zu den zuvor erwähnten Ansätzen beschreiben Oh und Oh (2011) keine Niveaus, sondern listen verschiedene elaborierte und somit wissenschaftliche Vorstellungen von den einzelnen Aspekten auf. Unter dem Aspekt *meanings of models* subsumieren sie beispielsweise die Vorstellungen **“a model is a representation of a target”** und **“a model serves as a ‘bridge’ or mediator connecting a theory and a phenomenon”** (S. 1124). Darüber hinaus formulieren Oh und Oh (2011) einen didaktischen Aspekt mit den Namen *uses of models in the science classroom*. Hier beschreiben sie, wie Modelle im naturwissenschaftlichen Unterricht eingesetzt werden, und dies verbinden sie mit der Forderung, im Unterricht Modelle als Medien und als Methode zu verwenden (Oh & Oh, 2011, S. 1124).

Sowohl Leisner-Bodenthin (2006) als auch Meisert (2008, 2009) unterteilen die Modellkompetenz ebenfalls theoretisch in verschiedene Komponenten (unter anderem deklaratives und prozedurales Wissen) und kombinieren die Dimensionen *Fachwissen* und *Erkenntnisgewinnung* miteinander. Für den Physikunterricht strukturierte Leisner-Bodenthin (2006) Modellkompetenz in *deklaratives Wissen*, *prozedurales Wissen* und *Selbstständigkeitsgrad*. Zum deklarativen Wissen zählt sie das *Modellverständnis* und das Wissen über den *Inhalt des Modells*. Prozedurales Wissen umfasst die *Anwendung des Inhalts* und die *Anwendung des Modellverständnisses*. „Der Selbstständigkeitsgrad äußert sich in der Umsetzbarkeit des deklarativen und prozeduralen Wissens in unterschiedlichen und unbekanntem Situationen“ (Leisner-Bodenthin, 2006, S. 93). Für diese verschiedenen Komponenten werden zwar unterschiedliche Vorstellungen festgehalten, jedoch keine Niveaus beschrieben. Darüber hinaus unterscheidet Leisner-Bodenthin (2006) zwischen einer domänenübergreifenden und einer domänenspezifischen Modellkompetenz. Während unter der ersten abstrakte Vorstellungen verstanden werden, werden mit der zweiten konkrete Vorstellungen verbunden.

Henze et al. (2007) beschreiben drei Ebenen (*learning of science*, *act of modelling* und *role and nature of models in science*), und auf dieser Grundlage differenziert Meisert (2008) drei Komponenten von Modellkompetenz: *Modell-*

wissen, Modellarbeit und *Modellverständnis*. Das *Modellwissen* bezieht sich auf Fachwissen und ist dem Aspekt *Inhalt des Modells* von Leisner-Bodenthin (2006) ähnlich. *Modellarbeit* schreibt sie der prozeduralen Komponente zu, allerdings erweitert sie im Gegensatz zu Leisner-Bodenthin (2006) den prozeduralen Aspekt um die eigene Entwicklung und die Weiterentwicklung von Modellen. Die prozedurale Komponente ordnet sie der Erkenntnisgewinnung zu. Der Aspekt *Modellverständnis* nimmt für Meisert (2008) eine übergeordnete Rolle ein. Für diesen Aspekt formuliert Meisert (2009) drei Niveaus, die auf die von Grosslight et al. (1991) beschriebenen Niveaus aufbauen. Darüber hinaus schildert sie vier Teilaspekte des Modellverständnisses, die besonders auf die Beziehung zwischen Modell und Original fokussieren (Meisert, 2009): *Modell-Original-Relation, Modellierer-Rolle, Entwicklungscharakter von Modellen* und *Funktion von Modellen als Mittel der Erkenntnisgewinnung*.

Neben diesen theoretisch hergeleiteten Ansätzen generierten Treagust et al. (2002) mit Hilfe einer Faktorenanalyse fünf Aspekte von Modellen: *models as multiple representations, models as exact replicas, models as explanatory tools, uses of scientific models* und *the changing nature of models*. Wie Oh und Oh (2011) beschreiben auch Treagust et al. (2002) keine Niveaus, sondern subsumieren unter jedem Aspekt unterschiedlich elaboriert formulierte Items.

Im Gegensatz zu den zuvor beschriebenen Ansätzen, die sich auf *Vorstellungen von Modellen* und damit auf die kognitive Ebene beziehen, fokussieren die Ansätze von Louca et al. (2011) und Schwarz et al. (2009) auf den *Umgang mit Modellen*. Louca et al. (2011) beispielsweise differenzieren in ihrem Strukturierungsvorschlag fünf Aspekte, die im Rahmen der Modellierung im Unterricht als Kriterien zur Beurteilung herangezogen werden können: *representation of physical objects, representation of object characteristics (physical entities), representation of interactions, representation of object behavior* und *accuracy of the phenomenon depiction*. Niveaus beschreiben Louca et al. (2011) für die einzelnen Aspekte nicht, jedoch sind den einzelnen Aspekten zwischen zwei und vier verschiedene Kategorien zugeordnet. So umfasst etwa der Aspekt *accuracy of the phenomenon depiction* die beiden Kategorien *non-valid* und *valid*.

Schwarz et al. (2009) integrieren in ihrem Rahmen den Umgang mit Modellen (*elements of the practice*), Vorstellungen von Modellen (*metamodelling knowledge*) und Kommunikationsaspekte (*sensemaking* und *communicating under-*

standing). Unter *metamodelling knowledge* fassen sie *nature of models*, *purpose of models* und *criteria for evaluating and revising models* zusammen (Schwarz et al., 2009). Unter der Dimension *elements of the practice* werden subsumiert: *constructing models*, *using models to explain and predict*, *evaluating models* und *revising models* (Schwarz et al., 2009). Die Integration dieser verschiedenen Komponenten spiegelt sich auch in den von Schwarz et al. (2009) formulierten vier Niveaus wider. So lautet das Niveau II für den Aspekt *using models to explain and predict* beispielsweise:

“Students construct and use a model to illustrate and explain how a phenomenon occurs, consistent with the evidence about the phenomenon. Students view models as a means of communicating their understanding of a phenomenon rather than a tool to support their own thinking” (Schwarz et al., 2009, S. 640).

Jedes Niveau besteht bei Schwarz et al. (2009) aus mehreren unterschiedlichen Komponenten. Beim Niveau II werden eine praktische und eine kognitive Komponente miteinander kombiniert. Ein solcher Ansatz wirft allerdings bei der Kompetenzmessung laut Klieme und Leutner (2006) sowie Weinert (1999, 2001) Probleme auf. Dies liegt daran, dass Lernende hinsichtlich dieser Komponenten unterschiedliche Fähigkeiten haben können. Damit wird es einerseits schwierig, einen Lernenden eindeutig einem spezifischen Niveau zuzuordnen, und andererseits können Zusammenhänge zwischen den Komponenten nicht näher untersucht werden. In einer aktuellen Veröffentlichung beschreiben Schwarz, Reiser, Acher, Kenyon und Fortus (2012) das Problem, dass es bei einer Anwendung des Rahmens nicht möglich war, Lernende eindeutig einem Niveau zuzuordnen. Dieser Befund unterstützt damit den Ansatz, solche Komponenten getrennt voneinander zu beschreiben und zu erfassen. Die Studie macht zudem deutlich, dass die empirische Überprüfung einer Strukturierung hilfreiche Informationen über die Adäquatheit eines solchen Rahmens liefert und damit unerlässlich ist.

Zusammenfassend ist festzuhalten, dass es verschiedene Ansätze gibt, die Vorstellungen von Modellen und der Modellbildung zu strukturieren (z. B. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Leisner-Bodenthin, 2006; Louca et al., 2011; Meisert, 2008; Oh & Oh, 2011; Schwarz et al., 2009; Treagust et al., 2002). Diese unterscheiden sich in ihrem jeweili-

gen Schwerpunkt (Fachwissen und/oder Erkenntnisgewinnung; kognitive Komponente von Modellkompetenz und/oder praktischer Umgang mit Modellen), in ihrer Dimensionierung in Aspekte sowie in ihrer Graduierung in verschiedene Ausprägungen (vgl. Grünkorn, Upmeier zu Belzen et al., 2014). Damit ist es für Lehrkräfte wie für Forschende schwierig zu entscheiden, auf welchen Ansatz sie sich bei ihren Fördermaßnahmen oder bei weiteren Untersuchungen berufen sollen. Zudem fehlt den meisten Strukturierungen eine empirische Überprüfung, die zeigen würde, inwiefern die Aspekte die Vorstellungen von Schülerinnen und Schülern hinreichend berücksichtigen, bzw., inwiefern sich gewisse Vorstellungen auch in anderen Zielgruppen empirisch beschreiben lassen. Demzufolge besteht ein Bedarf an einer Strukturierung, die wesentliche Ergebnisse dieser Studien zusammenfasst und empirisch geprüft ist.

2.3 Theoretische Struktur von Modellkompetenz

Die Grundlage dieser Arbeit bildet das theoriegeleitete Kompetenzstruktur- und Kompetenzniveaumodell (im Weiteren Kompetenzmodell) von Upmeier zu Belzen und Krüger (2010). Es gliedert sich in fünf Teilkompetenzen, die jeweils in drei Reflexionsniveaus differenziert sind: *Eigenschaften von Modellen*, *Alternative Modelle*, *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen* (Tab. 1).

Das Kompetenzmodell der Modellkompetenz bezieht sich auf die Domäne *Biologie*. Im Gegensatz zu Schwarz et al. (2009) sowie Leisner-Bodenthin (2006) und Meisert (2008, 2009), die verschiedene Komponenten bzw. Dimensionen von Modellkompetenz verknüpfen, konzentriert sich dieses Kompetenzmodell auf die Dimension *Erkenntnisgewinnung* und auf *kognitive* Strukturen. Diese kognitiven Strukturen sind grundlegend dafür, dass man Probleme in verschiedenen Situationen erfolgreich lösen kann (Fleischer, Koeppen, Kenk, Klieme & Leutner, 2013; Weinert, 1999, 2001).

Mit der Fokussierung auf kognitive Kompetenzen ist es möglich, die kognitive Komponente der Modellkompetenz zu erfassen und diese zu den nicht-kognitiven Fähigkeiten in Beziehung zu setzen. Letzteres ist wichtig, weil die nicht-kognitiven Fähigkeiten die kognitiven Fähigkeiten beeinflussen können (vgl. Weinert, 1999, 2001). Um diesen Aspekt Rechnung zu tragen, erweiterten Upmeier zu Belzen und Krüger (2010) ihre Definition von Modellkompetenz um eine volitionale Komponente. Sie definieren *Modellkompetenz* nun als die

Fähigkeit von Lernenden, „mit Modellen zweckbezogen Erkenntnisse gewinnen zu können und über Modelle mit Bezug auf ihren Zweck urteilen zu können, die Fähigkeiten, über den Prozess der Erkenntnisgewinnung durch Modelle und Modellierungen in der Biologie zu reflektieren sowie die Bereitschaft, diese Fähigkeiten in problemhaltigen Situationen anzuwenden“ (Upmeier zu Belzen & Krüger, 2010, S. 49). Damit schließen Upmeier zu Belzen und Krüger (2010) einerseits eine mediale und methodische Perspektive auf Modelle sowie ein konkretes und abstraktes Verständnis von Modellen ein, andererseits nehmen sie damit auch die volitionale Komponente in den Blick. Die vorliegende Forschungsarbeit fokussiert ausschließlich auf die kognitiven Kompetenzen.

Tab. 1: Niveaus und Teilkompetenzen des theoriegeleiteten Kompetenzmodells der Modellkompetenz (verändert nach Upmeier zu Belzen & Krüger, 2010)

Teilkompetenz	Komplexität		
	Niveau I	Niveau II	Niveau III
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	Modell als Modell von etwas durch neue Erkenntnisse oder zusätzliche Perspektiven revidieren	Modell für etwas aufgrund falsifizierter Hypothesen revidieren

Hinsichtlich der Entwicklung des Kompetenzmodells der Modellkompetenz basiert das Kompetenzmodell zunächst auf empirischen Studien (Tab. 2; Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003). Aus diesem Grund ist das Kompetenzmodell als ein deskriptives Kompetenzmodell zu verstehen (vgl. Schecker & Parchmann, 2006).

Dann ist das Kompetenzmodell auch als ein normatives Kompetenzmodell anzusehen (vgl. Schecker & Parchmann, 2006): In diesem wird mit der Berück-

sichtigung wissenschaftstheoretischer Sichtweisen (Gilbert, 1991; Mahr, 2008, 2009; Oh & Oh, 2011) und der fachspezifischen Bildungsstandards für das Fach Biologie normativ festgelegt, über welche Kompetenzen die Lernenden verfügen sollen. Im Kompetenzbereich **Erkenntnisgewinnung** (E) formuliert die Kultusministerkonferenz (KMK, 2005) fünf Standards (E 9 bis E 13), die sich auf ein Denken in Modellen und über Modelle beziehen. Demnach sollen die Schülerinnen und Schüler in der Regel mit Abschluss der zehnten Jahrgangsstufe dazu befähigt sein, Modelle einzusetzen, um komplexe Phänomene zu veranschaulichen, zu erklären und zu analysieren. Weiterhin sollen die Lernenden in der Lage sein, die Aussagekraft von Modellen zu reflektieren (KMK, 2005). Diese Standards finden sich im Kompetenzmodell der Modellkompetenz wieder (Tab. 2, S. 46f.).

Somit postulieren Upmeier zu Belzen und Krüger (2010) auf der Basis normativer Aspekte und deskriptiver Befunde ein theoretisch begründetes Kompetenzmodell mit drei Niveaus pro Teilkompetenz (vgl. S. 42ff. dieser Arbeit) und fünf Teilkompetenzen (vgl. S. 44ff. dieser Arbeit). Damit ist eine Voraussetzung für eine Operationalisierung und eine Erfassung von Modellkompetenz geschaffen (vgl. Klieme & Leutner, 2006). Aufgrund des hypothetischen und in Teilen normativen Charakters des Kompetenzmodells bleibt zu prüfen, ob es die Perspektiven von Schülerinnen und Schülern hinreichend reflektiert.

2.3.1 Graduierung in Niveaus

Upmeier zu Belzen und Krüger (2010) beschreiben drei Niveaus pro Teilkompetenz, die durch unterschiedlich elaborierte epistemologische Positionen charakterisiert sind (Tab. 1). Diese Niveaus orientieren sich zwar grundlegend an den drei generellen Niveaus von Grosslight et al. (1991), berücksichtigen aber zusätzlich die Ergebnisse der Studie von Justi und Gilbert (2003). So unterstützten Justi und Gilbert (2003) nach Abschluss ihrer Untersuchung den Ansatz, pro Teilkompetenz jeweils ein Niveau zu definieren. Auch Crawford und Cullin (2005) formulieren in ihrem Strukturierungsansatz bereits für jeden Aspekt zwischen drei und vier verschiedene Niveaus. Diese Befunde verknüpfen Upmeier zu Belzen und Krüger (2010) mit den drei Perspektiven auf Modelle von Mahr (2008, 2009), die eine Interpretationsebene für die drei Niveaus bereitstellen: die Modellobjektebene, die Herstellungsperspektive (Modell **von** etwas) und die Anwendungsperspektive (Modell **für** etwas).

Die von Upmeier zu Belzen und Krüger (2010) beschriebenen Niveaus reflektieren damit – allgemein gesprochen – die wissenschaftstheoretische Sicht auf Modelle als *Medien* zur Veranschaulichung sowie als *Methode* zur Erkenntnisgewinnung (Gilbert, 1991; Mahr, 2008, 2009; Oh & Oh, 2011): Auf *Niveau I* wird zunächst das Modellobjekt an sich betrachtet. Damit liegt der Fokus auf der Darstellungsform des Modells, bei der das Original unbeachtet bleibt. Auf *Niveau II* wird das Modell als Modell *von* etwas (Herstellungsperspektive) wahrgenommen und damit die Beziehung von Modell und Original näher betrachtet. Die methodische Sicht auf Modelle als Modelle *für* etwas (Anwendungsperspektive) beschreibt Niveau III. Hier werden Modelle als Mittel der Erkenntnisgewinnung verstanden, mit denen Ideen hinsichtlich des Originals getestet werden (Gilbert, 1991; Mahr, 2008, 2009; Oh & Oh, 2011). Die allgemeine Charakterisierung der drei Niveaus zeigt sich jedoch nicht bei allen Teilkompetenzen. Eine Ausnahme stellen die Teilkompetenzen *Eigenschaften von Modellen* und *Zweck von Modellen* dar. Bei beiden Teilkompetenzen steht die Beziehung von Modell und Original im Fokus, so dass eine isolierte Betrachtung des Modellobjektes (bei den übrigen Teilkompetenzen: Niveau I) nicht möglich ist (vgl. Upmeier zu Belzen & Krüger, 2010 S. 50). Die Niveaus unterscheiden sich hinsichtlich dieser Teilkompetenzen durch unterschiedlich elaborierte Sichtweisen der Beziehung von Modell und Original (Tab. 1). Bei *Eigenschaften von Modellen* äußert sich dies beispielsweise darin, dass das Modell auf Niveau I als Kopie, auf Niveau II als idealisierte Repräsentation und auf Niveau III als theoretische Rekonstruktion wahrgenommen wird (Tab. 1).

Die drei Niveaus für jede Teilkompetenz verstehen Upmeier zu Belzen und Krüger (2010) als hierarchische Ausprägungen mit steigender Komplexität. Aktuelle Ergebnisse aus den Studien von Krell (2012) und Terzer (2013) unterstützen diese Annahme. Sie konnten zeigen, dass sich die Niveaus signifikant unterscheiden und mit einem steigenden Schwierigkeitsgrad einhergehen. Inwiefern diese Niveaus darüber hinaus auch als Entwicklungsstufen zu interpretieren sind, ist noch zu prüfen (Patzke, Krüger & Upmeier zu Belzen, *in Druck*). Ebenso ist zu prüfen, ob sich in den empirischen Daten zusätzliche Niveaus für bestimmte Teilkompetenzen zeigen. Diese Klärung ist nicht zuletzt auch deshalb vonnöten, weil Crawford und Cullin (2005) für die Aspekte *purpose of models*, *changing a model* und *multiple models for the same thing* vier Niveaus definiert haben.

Weiterhin ist festzustellen, dass Kompetenzmodelle mit Blick auf die Beschreibung der Niveaus häufig abstrakt formuliert sind (Klieme, 2004). Dies zeigt sich auch beim Kompetenzmodell der Modellkompetenz (Upmeier zu Belzen & Krüger, 2010). Wie beispielsweise das Niveau I der Teilkompetenz **Ändern von Modellen, Mängel am Modellobjekt beheben**, in konkretere Perspektiven differenziert wird, ist zwar kein Bestandteil des Kompetenzmodells selbst, jedoch ein wichtiger Ansatzpunkt für Fördermaßnahmen. Aus diesem Grund ist eine Ausdifferenzierung der einzelnen Niveaus ein Ziel der vorliegenden Arbeit (vgl. Grünkorn, Upmeier zu Belzen et al., 2014).

2.3.2 Dimensionierung in Teilkompetenzen

Das Kompetenzmodell der Modellkompetenz ist in fünf Teilkompetenzen unterteilt: *Eigenschaften von Modellen, Alternative Modelle, Zweck von Modellen, Testen von Modellen* und *Ändern von Modellen* (Tab. 2). Die Teilkompetenzen *Eigenschaften von Modellen* und *Alternative Modelle* beschreiben individuelle ontologische und epistemologische Konzepte von Modellen und wurden zur Dimension *Kenntnisse über Modelle* zusammengefasst (vgl. Günther, 2006; Mayer, 2007; Upmeier zu Belzen & Krüger, 2010). Die Teilkompetenzen *Zweck von Modellen, Testen von Modellen* und *Ändern von Modellen* beziehen sich auf kognitive Fähigkeiten im Modellbildungsprozess und strukturierten die Dimension *Modellbildung* (vgl. Justi & Gilbert, 2002, 2006; Upmeier zu Belzen & Krüger, 2010). Beide Dimensionen beruhen auf übergeordneten Zielen der naturwissenschaftlichen Bildung, die von Henze et al. (2007), Hodson (1992), Mayer (2007) und Meisert (2008) aufgeführt werden. So kann die Dimension *Kenntnisse über Modelle* den Zielen *learning about science* (Hodson, 1992), *role and nature of models in science* (Henze et al., 2007), *Modellverständnis* (Meisert, 2008) und *Wissenschaftsverständnis* (Mayer, 2007) zugeordnet werden. Die Dimension *Modellbildung* ist dagegen mit den Zielen *doing science* (Hodson, 1992), *act of modelling* (Henze et al., 2007), *Modellarbeit* (Meisert, 2008) und *wissenschaftliches Denken* sowie *manuelle Fertigkeiten* (Mayer, 2007) vereinbar. In Studien zur Dimensionalität des Kompetenzmodells der Modellkompetenz von Krell (2013; Krell & Krüger, 2011; Krell, Upmeier zu Belzen & Krüger, 2012a) sowie Terzer (2013) konnte allerdings die zweidimensionale Struktur nicht bestätigt werden. Aus diesem Grund wird im Weiteren die Differenzie-

rung in die zwei Dimensionen *Kenntnisse über Modelle* und *Modellbildung* nicht näher in den Blick genommen.

Die Herleitung der Teilkompetenzen des Kompetenzmodells basiert hauptsächlich auf den Strukturierungsansätzen von Crawford und Cullin (2005), Grosslight et al. (1991) sowie Justi und Gilbert (2003). Tabelle 2 zeigt in der Übersicht, welche Aspekte grundlegend für die Entwicklung dieser Teilkompetenzen waren (grau hinterlegt). Zudem werden auch die in weiteren Studien generierten Aspekte aufgeführt, die ebenfalls von den Teilkompetenzen des Kompetenzmodells abgedeckt werden (nicht hinterlegt, Tab. 2). Aspekte, die primär den Handlungen zuzuordnen sind, sind bei der Beschreibung der Teilkompetenzen nur hinsichtlich ihrer kognitiven Komponente berücksichtigt worden. Dies zeigt sich beispielsweise beim Aspekt *designing and creating models* von Grosslight et al. (1991). Grosslight et al. (1991) unterteilten den Aspekt in verschiedene Kategorien, wie *depends on the view* oder *depends on what is major/minor, basics/details, little things/big thinks, main thinks*. Diese Kategorien lassen sich auf die Teilkompetenz *Alternative Modelle* beziehen.

Im Hinblick auf die für das Fach Biologie formulierten Standards von Modellen (E 9 bis E 13) zeigt die Zuordnung der Standards, dass sich diese ausschließlich auf die Teilkompetenzen *Zweck von Modellen* und *Testen von Modellen* (KMK, 2005) konzentrieren (Tab. 2). Andere Teilkompetenzen des Kompetenzmodells werden in den Standards nicht explizit angesprochen. Das ist nicht ungewöhnlich (Klieme, Maag Merki et al., 2007). Der Grund hierfür liegt in der Fokussierung der Standards auf Kernbereiche. Kompetenzmodelle übernehmen die Aufgabe, Standards auszugestalten und alle notwendigen Aspekte zu benennen, die zur Entwicklung dieser konzentrierten Kernbereiche beitragen. So postulieren Upmeier zu Belzen und Krüger (2010), dass ein elaboriertes Wissenschaftsverständnis die Entwicklung einer medialen und methodischen Perspektive in den Teilkompetenzen *Zweck* und *Testen von Modellen* unterstützen kann. Dieses Wissenschaftsverständnis spiegelt sich in den Teilkompetenzen *Eigenschaften von Modellen* und *Alternative Modelle* wider, weshalb es wichtig ist, auch diese Teilkompetenzen zu fördern. Weiterhin kann die Aussagekraft von Modellen nur dann beurteilt werden (E 13, KMK, 2005), wenn auch das *Ändern von Modellen* als Folge des *Testens von Modellen* entwickelt wird. Auf diese Weise ist es möglich, dass Kompetenzmodelle mehr Aspekte berücksichtigen als die in den Standards genannten (Klieme, Maag Merki et al., 2007).

Tab. 2: Strukturierung der aus der Literatur genannten Aspekte von Modellen zu Teilkompetenzen (verändert nach Krell, Upmeier zu Belzen & Krüger, 2014a sowie Upmeier zu Belzen & Krüger, 2010)

Quelle	Teilkompetenzen des Kompetenzmodells der Modellkompetenz				
	Eigenschaften von Modellen	Alternative Modelle	Zweck von Modellen	Testen von Modellen	Ändern von Modellen
Grosslight et al. (1991)	kinds of models (designing and creating models)	multiple models for the same thing (designing and creating models)	purpose models (designing and creating models)	purpose of models – testing (designing and creating models)	changing a model (designing and creating models)
Treagust et al. (2002)	models as exact replicas	models as multiple representations	models as explanatory tools uses of scientific models	–	the changing nature of models
Justi & Gilbert (2003)	nature entities	uniqueness	use predictions	accreditation	time
Crawford & Cullin (2005)	–	multiple models for the same thing (designing and creating models)	purpose of models	validating/testing models	changing a model (designing and creating models)
KMK (2005) ¹	–	–	E 9; E 10; E 11; E 12	E 13	–

Anmerkungen. Die grau hinterlegten Zeilen markieren die Studien, die grundlegend für die Entwicklung der Teilkompetenzen sind. Die nicht hinterlegten Zeilen kennzeichnen Studien, die zwar nicht in die Entwicklung der Teilkompetenzen einfließen, aber trotzdem abgedeckt werden. Aspekte, die in Klammern gesetzt sind, beziehen sich auf Handlungen und sind nur indirekt mit Blick auf die kognitive Komponente den Teilkompetenzen zugeordnet. Die Quellen sind nach Erscheinungsjahr geordnet.

¹ Standards im Kompetenzbereich **Erkenntnisgewinnung**, die sich auf Modelle beziehen lassen (KMK, 2005): E 9: Schülerinnen und Schüler wenden Modelle zur Veranschaulichung von Struktur und Funktion an. E 10: Schülerinnen und Schüler analysieren Wechselwirkungen mit Hilfe von Modellen. E 11: Schülerinnen und Schüler beschreiben Speicherung und Weitergabe genetischer Information auch unter Anwendung geeigneter Modelle. E 12: Schülerinnen und Schüler erklären dynamische Prozesse in Ökosystemen mithilfe von Modellvorstellungen. E 13: Schülerinnen und Schüler beurteilen die Aussagekraft eines Modells.

Fortsetzung

Tab. 2 (Fortsetzung)

Quelle	Teilkompetenzen des Kompetenzmodells der Modellkompetenz				
	Eigenschaften von Modellen	Alternative Modelle	Zweck von Modellen	Testen von Modellen	Ändern von Modellen
Meisert ² (2009)	Modell-Original-Relation	–	Entwicklungscharakter von Modellen	Entwicklungscharakter von Modellen	Entwicklungscharakter von Modellen
	Modellierer-Rolle		Funktion von Modellen als Mittel der Erkenntnisgewinnung	Funktion von Modellen als Mittel der Erkenntnisgewinnung	Funktion von Modellen als Mittel der Erkenntnisgewinnung
Schwarz et al. (2009)	nature of models	–	purpose of models (uses models to explain and predict)	criteria for evaluating and revising models (evaluating models)	criteria for evaluating and revising models (revising models)
Oh & Oh (2011)	meaning of models	multiplicity of scientific models	purposes of modelling (uses of models in the science classroom)	changes in scientific models	changes in scientific models

Anmerkungen. Die nicht hinterlegten Zeilen kennzeichnen Studien, die zwar in die Entwicklung der Teilkompetenzen nicht einfließen, aber trotzdem abgedeckt werden. Aspekte, die in Klammern gesetzt sind, beziehen sich auf Handlungen und sind nur indirekt mit Blick auf die kognitive Komponente den Teilkompetenzen zugeordnet. Die Quellen sind nach Erscheinungsjahr geordnet.

² Für die Dimension **Modellarbeit** sind bislang keine formuliert, weswegen nur Aspekte der Dimension **Modellverständnis** beschrieben sind (Meisert, 2008, 2009).

Vor diesem Hintergrund klassifizieren Upmeyer zu Belzen und Krüger (2010) die folgenden Teilkompetenzen mit ihren drei Reflexionsniveaus (Tab. 1): Die Teilkompetenz **Eigenschaften von Modellen** befasst sich mit der Beziehung von Modell und Original (vgl. Grosslight et al., 1991; Justi & Gilbert, 2003; Tab. 2). Diese Beziehung kann unterschiedlich charakterisiert werden (Tab. 1). Auf Niveau I werden Modelle als Kopien von etwas und auf Niveau II als idealisierte Repräsentationen von etwas verstanden. Die elaborierteste Sicht auf Modelle spiegelt Niveau III wider, in dem Modelle als theoretische Rekonstruktionen wahrgenommen werden.

Die Teilkompetenz **Alternative Modelle** bildet ab, dass ein Original durch verschiedene Modelle repräsentiert sein kann (vgl. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Tab. 2). Diese alternativen Modelle eines Originals lassen sich unterschiedlich begründen (Tab. 1). Auf Ni-

veau I werden Unterschiede in den Modellobjekten beschrieben. Auf Niveau II beruhen alternative Modelle auf der Komplexität des Originals, das die Konstruktion verschiedener Modelle ermöglicht. Niveau III nimmt alternative Modelle eines Originals als Modelle für etwas in den Blick. Dies äußert sich darin, dass die Existenz alternativer Modelle durch verschiedene Hypothesen über das Original erklärt wird.

Beim **Zweck von Modellen** wird der unterschiedliche Nutzen von Modellen im Modellierungsprozess betrachtet (vgl. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Tab. 2). Dabei lassen sich drei verschiedene Zwecke voneinander differenzieren (Tab. 1): Modelle können dazu eingesetzt werden, um Merkmale des Originals zu beschreiben (Niveau I) oder um bekannte Zusammenhänge mit dem Original wiederzuerkennen und zu erklären (Niveau II). Auf Niveau III sind Modelle als Instrumente zu verstehen, mit denen neue Erkenntnisse gewonnen und Vorhersagen über das Original getroffen werden.

Beim **Testen von Modellen** wird beschrieben, wie Modelle überprüft werden können (vgl. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Tab. 2). Auf Niveau I wird zunächst das Modellobjekt getestet, beispielsweise indem die Widerstandsfähigkeit des verwendeten Materials geprüft wird. Auf Niveau II wird das Modell mit dem Original verglichen. Niveau III beschreibt die elaborierteste Sicht auf Modelle, in der mit dem Modell eine Hypothese über das Original untersucht wird (Tab. 1).

Das Ergebnis einer Testung kann zu Veränderungen am Modell führen (vgl. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Tab. 2). Die Gründe für ein **Ändern von Modellen** sind vielfältig (Tab. 1): Auf Niveau I liegt der Grund in Mängeln am Modellobjekt. Auf Niveau II führen neue Erkenntnisse über das Original dazu, dass ein Modell verändert wird. Liegt der Grund für eine Veränderung des Modells darin, dass eine Hypothese, die dem Modell zugrunde liegt, bei der Anwendung des Modells falsifiziert wurde, wird eine methodische Sicht auf Modelle beschrieben (Niveau III).

3 Zielsetzungen der Arbeit

Sowohl auf nationaler (Leisner-Bodenthin, 2006; Meisert, 2008, 2009) als auch internationaler (u. a. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Louca et al., 2011; Oh & Oh, 2011; Schwarz et al., 2009, 2012) Ebene gibt es unterschiedliche Ansätze, wie Vorstellungen von Modellen und von dem Modellbildungsprozess strukturiert und graduiert werden. Diese Strukturierungsansätze wurden – mit Ausnahme der Strukturierung von Schwarz et al. (2012) – bisher nicht auf einer breiten Datengrundlage von Lernenden aus der Zielgruppe geprüft. Deshalb kann gegenwärtig nur angenommen werden, dass diese Ansätze die Perspektiven der Lernenden hinreichend reflektieren. Darüber hinaus erschwert es die Unterschiedlichkeit dieser Strukturierungsansätze entscheidend, Studienergebnisse zuverlässig miteinander zu vergleichen (Hartig, 2008; Klieme & Hartig, 2007; Klieme et al., 2008; Klieme, Maag Merki et al., 2007).

Daher leisten Upmeyer zu Belzen und Krüger (2010) mit ihrem theoretisch hergeleiteten Kompetenzmodell einen wichtigen Beitrag zu diesem Forschungsbereich: Sie postulieren ein Kompetenzstruktur- und Kompetenzniveaumodell, das einerseits unterschiedliche Ansätze der Strukturierung und Graduierung der Vorstellungen von Modellen und vom Modellbildungsprozess aufgreift und integriert (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003). Andererseits beziehen sie wissenschaftstheoretische Literatur zu Modellen (u. a. Gilbert, 1991; Mahr, 2009) ein und nehmen Bezug auf die Bildungsstandards im Kompetenzbereich Erkenntnisgewinnung des Unterrichtsfachs Biologie (E 9 bis E 13, KMK, 2005).

Eine empirische Überprüfung dieses theoretisch begründeten Kompetenzmodells steht jedoch aus. Hierzu eignen sich neben geschlossenen Antwortformaten auch offene Formate zur Operationalisierung. Letztere erlauben Einblicke in die Perspektiven von Schülerinnen und Schülern und ermöglichen eine inhaltliche Überprüfung von Kompetenzmodellen (vgl. Eggert & Bögeholz, 2010; Grube, 2011; Rost, 2004). Dementsprechend stehen die folgenden zwei Ziele im Zentrum der vorliegenden Arbeit (Abb. 4):

1. Das Kompetenzmodell der Modellkompetenz soll mit Aufgaben im offenen Antwortformat für die Domäne **Biologie** operationalisiert werden (vgl. S. 53ff. dieser Arbeit) und darauf aufbauend:

2. Die Modellkompetenz von Schülerinnen und Schülern der Sekundarstufe I analysiert und beschrieben werden (vgl. S. 85ff. dieser Arbeit).

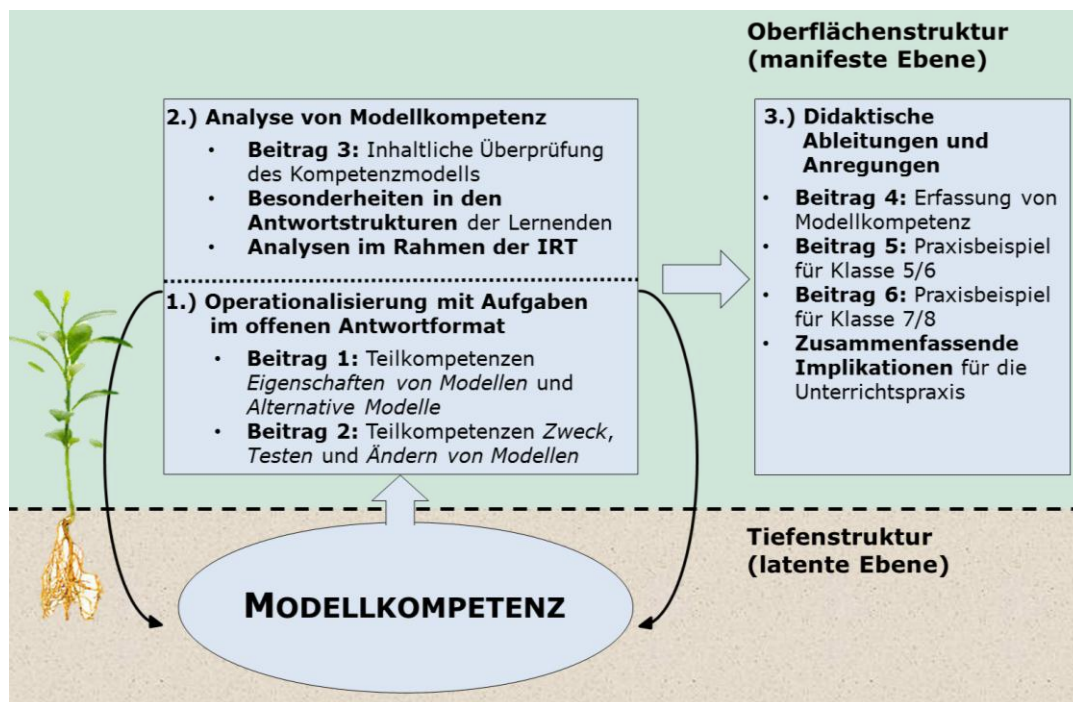


Abb. 4: Das Forschungsdesign und die drei Ziele des Dissertationsprojekts (Grafik nach Meyer, 2012)

Das Forschungsdesign dieser Arbeit (Abb. 4) baut auf den Vorgehensweisen von Studien auf, die analoge Ziele verfolgten (u. a. Eggert, 2008; Grube, 2011; Schmiemann, 2010). **Beitrag 1** (Grünkorn et al., 2011) beschreibt die Entwicklung, Überprüfung und Selektion von Aufgaben im offenen Antwortformat für die Teilkompetenzen *Eigenschaften von Modellen* und *Alternative Modelle* und **Beitrag 2** (Grünkorn & Krüger, 2012) für die Teilkompetenzen *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen*. Als Ergebnis dieses Operationalisierungs- und Überprüfungsprozesses steht ein Leistungstest mit 15 Aufgaben im offenen Antwortformat zur Verfügung.

Dieser Leistungstest wird für die Studie zur Beschreibung der Modellkompetenz von Schülerinnen und Schülern der Sekundarstufe I eingesetzt. In dieser Studie wird geprüft, inwiefern das Kompetenzmodell der Modellkompetenz die inhaltlichen Aspekte der einzelnen Teilkompetenzen angemessen widerspiegelt. Dafür werden verschiedene Perspektiven auf Modelle und auf den Modellbildungsprozess sowie deren Häufigkeiten bei Schülerinnen und Schülern identifiziert.

ziert. Die Ergebnisse dieser inhaltlichen Überprüfung des Kompetenzmodells sind in **Beitrag 3** (Grünkorn, Upmeier zu Belzen et al., 2014) dargestellt.

Darüber hinaus steht die Analyse von Besonderheiten bei den Antworten der Lernenden im Mittelpunkt (Abb. 4). Sie gibt Einblicke in die Antwortstrukturen von Lernenden, ermöglicht dadurch ein vertieftes Verständnis und liefert Erklärungsansätze für die Perspektiven von Schülerinnen und Schülern auf Modelle und auf den Modellbildungsprozess (vgl. S. 104ff. dieser Arbeit).

Ergänzend zu diesem qualitativen Zugang werden quantitative Analysen im Rahmen der *Item-Response-Theorie* durchgeführt (Abb. 4). Die hier entwickelten Aufgaben im offenen Antwortformat sind Bestandteil eines Aufgabenpools mit MC-Items (Terzer, 2013), die Patzke et al. (*in Druck*) in ihrer Studie zur Entwicklung von Modellkompetenz im Biologieunterricht einsetzen. In Vorbereitung darauf wird geprüft, ob die 15 Aufgaben im offenen Antwortformat die notwendigen Gütekriterien für einen probabilistischen Ansatz erfüllen (vgl. S. 111ff. dieser Arbeit).

Aus den in dieser Arbeit durchgeführten Analysen insgesamt werden didaktische Empfehlungen und Anregungen abgeleitet. In der Literatur sind bereits einige Hinweise beschrieben, die Lehrkräfte für die Entwicklung von Fördermaßnahmen heranziehen können (u. a. Grosslight et al., 1991; Justi & Gilbert, 2003; Leisner-Bodenthin, 2006; Louca et al., 2011; Schwarz et al., 2009, 2012). Bislang fehlen allerdings schulalltagstaugliche Instrumente und konkrete Anregungen für einen (modell-)kompetenzorientierten Unterricht. Daraus ergibt sich das dritte Ziel dieser Forschungsarbeit (Abb. 4):

3. Didaktische Anregungen für einen (modell-)kompetenzorientierten Unterricht sollen aus den empirischen Analysen abgeleitet werden (vgl. S. 131ff. dieser Arbeit).

Beitrag 4 (Grünkorn, Lotz et al., 2014) stellt hierzu zentrale Perspektiven von Schülerinnen und Schülern auf Modelle und auf den Modellbildungsprozess vor, die in der Studie von Grünkorn, Upmeier zu Belzen et al. (2014) herausgearbeitet wurden und mit denen Lehrkräfte im Unterricht konfrontiert werden können. Im Sinne eines (modell-)kompetenzorientierten Unterrichtens präsentieren Grünkorn, Lotz et al. (2014) ein erprobtes Kompetenzraster, das Schülerinnen und Schülern lernbegleitend Rückmeldungen zu ihrem Kompetenz-

stand gibt und sie auf diese Weise dabei unterstützt, ihren Lernprozess zielgerichtet zu strukturieren. Ferner wird gezeigt, wie durch den Einsatz eines Diagnosebogens die Modellkompetenz von Lernenden auf schulalltagstaugliche Art und Weise eingeschätzt und mithilfe des Kategoriensystems der Modellkompetenz eingeordnet werden kann.

Neben einer Einschätzung des Kompetenzstandes von Schülerinnen und Schülern stellt sich die Frage, auf welche Weise die beschriebenen Implikationen aus der Forschung in den Unterricht integriert werden können. **Beitrag 5** (Grünkorn & Hanauer, 2013a, 2013b) und **Beitrag 6** (Grünkorn & Fleige, 2012) beschreiben zwei Unterrichtskonzepte, die Anregungen dafür geben, wie Modellkompetenz im Biologieunterricht gefördert werden kann. Beitrag 5 präsentiert ein Unterrichtsbeispiel zum Thema „**Untersuchung der Schleimschicht von Fischen**“, dessen Zielgruppe Schülerinnen und Schüler der fünften und sechsten Jahrgangsstufe (im Folgenden: Jgst.) sind. Beitrag 6 skizziert einen **Unterrichtsentwurf zum Thema „Bau und Funktion der Fischhaut**“, der für Schülerinnen und Schüler der siebten und achten Jahrgangsstufe geeignet ist. Darüber hinaus werden aus den Analysen weitere Anregungen und Empfehlungen für die Unterrichtspraxis abgeleitet und zusammengetragen (Abb. 4). Dabei wird zusammengefasst, welche Kenntnisse für Lehrkräfte relevant sind, wollen sie die Modellkompetenz von Schülerinnen und Schülern fördern, und welche Aspekte für eine effektive Förderung von Modellkompetenz einzubeziehen sind (vgl. S. 137ff. dieser Arbeit).

4 Operationalisierung des Kompetenzmodells der Modellkompetenz

Kompetenzen sind als kontextspezifische Leistungsdispositionen definiert (Klieme & Leutner, 2006), die in der Tiefenstruktur (latente Ebene) zu finden und somit nicht direkt beobachtbar und messbar sind (Abb. 5).

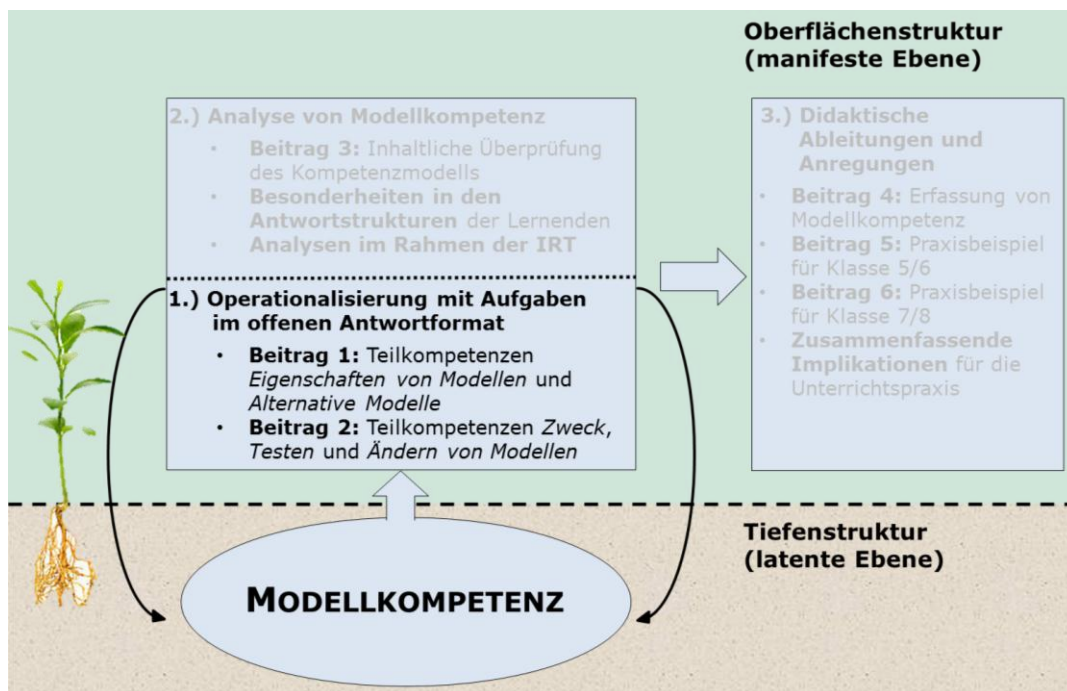


Abb. 5: Erstes Ziel der Forschungsarbeit: Operationalisierung des Kompetenzmodells der Modellkompetenz (Grafik nach Meyer, 2012)

Um Modellkompetenz erfassen und damit überprüfen zu können, inwiefern das Kompetenzmodell der Modellkompetenz die Perspektiven von Schülerinnen und Schülern auf Modelle und auf die Modellbildung hinreichend und angemessen beschreibt, ist ein zuverlässiger Test notwendig (vgl. Klieme, Avenarius et al. 2007; Schecker & Parchmann, 2006). Dieser Test muss Aufgaben enthalten, die als Indikatoren von Modellkompetenz interpretiert werden können (vgl. Hartig, Frey & Jude, 2012; Kane 2001, 2013). Erst dann ist es möglich, von der Performanz – dem Lösen von Testaufgaben (manifeste Ebene) – auf die Kompetenz der Lernenden (latente Ebene) zu schließen (Abb. 5). Bei der systematischen Entwicklung eines solchen Tests werden die einzelnen Schritte klar und nachvollziehbar beschrieben. Dadurch lässt sich die Qualität dieses Prozesses transparent machen. Neben allgemeinen Vorschlägen zu den einzel-

nen Schritten der Testkonstruktion (Bühner, 2010; Lienert & Raatz, 1998; Neuhaus & Braun, 2007; Rost, 2004) beschreiben Terzer et al. (2013) eine systematische Testentwicklung für MC-Items zur Modellkompetenz. Analog dazu wurde in der vorliegenden Arbeit die Testentwicklung für Aufgaben im offenen Antwortformat konzipiert, welche die entsprechenden Schritte für dieses Format konzeptualisiert (vgl. S. 54ff. dieser Arbeit). Auf der Grundlage dieser Vorgehensweise wurden die Teilkompetenzen *Eigenschaften von Modellen* und *Alternative Modelle* (Grünkorn et al., 2011) sowie *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen* (Grünkorn & Krüger, 2012) operationalisiert.

4.1 Methode der Testkonstruktion

Die Operationalisierung von Modellkompetenz erfolgt in fünf Schritten: (1) theoretische Fundierung, (2) Testkonzeption, (3) Aufgabenformat und auswertung, (4) Aufgabenentwicklung und (5) Aufgabenüberprüfung und -selektion.

4.1.1 Theoretische Fundierung

Für die Operationalisierung von Modellkompetenz sind sowohl die Kompetenzdefinition von Klieme und Leutner (2006, vgl. S. 15ff. dieser Arbeit) als auch das Kompetenzmodell der Modellkompetenz von Upmeyer zu Belzen und Krüger (2010) grundlegend. Die vorliegende Untersuchung bezieht sich darauf und greift daher nicht alle Facetten der Kompetenzdefinition auf, sondern nimmt die *kognitive Komponente* von Modellkompetenz in den Blick und klammert die motivationale, die volitionale und die soziale Komponente von Modellkompetenz aus.

Weiterhin bezieht sich die Kompetenzdefinition auf die *Kontextspezifität* und auf *Anforderungssituationen* (Klieme & Leutner, 2006). Dementsprechend lässt sich Modellkompetenz nicht mit einzelnen, isolierten Leistungen adäquat rekonstruieren, sondern es werden hierfür hinreichend viele verschiedene konkrete Situationen bzw. Lernkontexte benötigt. Diese Lernkontexte müssen zudem unterschiedliche Anforderungen – beispielsweise in Form von Niveaus – angemessen widerspiegeln (Hartig, 2008; Klieme & Hartig, 2007). Das bedeutet, dass die gewählten Kontexte einerseits Antworten der Lernenden auf den drei Niveaus der jeweiligen Teilkompetenz zulassen müssen, andererseits re-

suliert aus der Kontextspezifität von Kompetenz, dass die zu entwickelnden Aufgaben in verschiedene konkrete Situationen einzubetten sind (vgl. S. 66ff. dieser Arbeit).

Über diese Kontextspezifität hinaus ist der verwendete Kompetenzbegriff **domänenspezifisch**. Demgemäß werden Kompetenzen in einer bestimmten Domäne erworben und angewandt. Infolgedessen sind Kompetenzen nicht ohne Weiteres mit Bezug auf andere Domänen zu verallgemeinern (Klieme & Steinert, 2004). Das Kompetenzmodell der Modellkompetenz und die gewählten Kontexte beziehen sich auf die Domäne Biologie. Schlussfolgerungen aus den Ergebnissen der Studien sind dementsprechend auf das Fach Biologie begrenzt und können nur hypothetisch auf andere Domänen oder Fächer, wie Chemie und Physik, übertragen werden.

Neben der Kompetenzdefinition dient das Kompetenzmodell der Modellkompetenz als theoretischer Rahmen (vgl. S. 40ff. dieser Arbeit). Diese Strukturierung erlaubt es, Modellkompetenz mit Aufgaben systematisch zu operationalisieren (vgl. Klieme & Hartig, 2007; Lienert & Raatz, 1998).

4.1.2 Testkonzeption

Im Rahmen der Testkonzeption stellt sich zunächst die Frage nach dem **Einsatzbereich** bzw. den **Geltungsbereich** des Tests. Je nach Anwendungskontext, beispielsweise Überprüfung von Kompetenzmodellen, Systemmonitoring, Schulevaluation oder Individualdiagnostik, sind an einen Test unterschiedliche Bedingungen und Einschränkungen geknüpft (Jonkisz, Moosbrugger & Brandt, 2012; Kauertz, 2014; Klieme, Avenarius et al., 2007; Leutner et al., 2007; Neuhaus & Braun, 2007). Das Ziel dieser Forschungsarbeit war es, das Kompetenzmodell der Modellkompetenz empirisch zu überprüfen. Dafür war ein Test zu konstruieren, der mit Aufgaben das Konstrukt Modellkompetenz hinreichend repräsentiert (Hartig et al., 2012; Klieme, Avenarius et al., 2007). Im Gegensatz zur Individualdiagnostik, bei der in der Regel eine hohe Messgenauigkeit (Reliabilität) notwendig ist, sind für die empirische Überprüfung von Kompetenzmodellen auf Populationsebene auch niedrigere Messgenauigkeiten akzeptabel (Klieme, Avenarius et al., 2007; Leutner et al., 2007; Terzer et al., 2013).

Neben der Entscheidung über den Geltungsbereich ist die **Zielgruppe** zu bestimmen, über die mit dem Test Aussagen getroffen werden sollen (Jonkisz et

al., 2012). Das Kompetenzmodell der Modellkompetenz orientiert sich unter anderem an den Bildungsstandards für das Unterrichtsfach Biologie (KMK, 2005). Dort werden Regelstandards beschrieben, die Schülerinnen und Schüler mit Ende des mittleren Schulabschlusses erfüllen können sollen (KMK, 2005). Aus diesem Grund wurde der entwickelte Test für Lernende in Realschulen⁴ und Gymnasien (7. bis 10. Jahrgangsstufe) konzipiert, und die Stichprobe für die Erprobung des Tests aus dieser Zielgruppe gezogen.

Weitere Fragen der Testkonzeption beziehen sich darauf, was erfasst werden soll (Leistung oder Persönlichkeitsmerkmale) und mit welcher **Testart** dies zu bewältigen ist. Mit Blick auf das Rahmenkonzept wissenschaftsmethodischer Kompetenzen sind Modelle und Modellierungsprozesse den Dimensionen **Charakteristika der Naturwissenschaften (nature of science)** und **wissenschaftliche Untersuchungen (scientific inquiry)** zuzuordnen (Mayer, 2007). Dementsprechend ist für die Erfassung solcher Kompetenzen zunächst ein **Performance Test** einem schriftlichen Leistungstest vorzuziehen, weil dieser manuelle Handlungen und ein exploratives Vorgehen misst. In einem Performance Test werden den Lernenden verschiedene Experimentalsituationen angeboten, für deren Bewältigung unterschiedliche Fähigkeiten benötigt werden (Shavelson, Baxter & Gao, 1993). Orsenne und Upmeyer zu Belzen (2012) nutzen in einem parallelen Forschungsprojekt dieses Format, um die manuellen Handlungen im Modellierungsprozess zu untersuchen. Im Mittelpunkt des vorliegenden Forschungsprojekts stehen allerdings die kognitive Komponente der Modellkompetenz und die empirische Überprüfung des Kompetenzmodells an einer großen Anzahl von Lernenden aus verschiedenen Jahrgangsstufen. Aus diesen Gründen ist ein schriftlicher Leistungs-Powertest geeigneter als ein Performance Test (Kauertz et al., 2010). Dieser Leistungstest ist so zu konzipieren, dass die Zeit für eine vollständige Bearbeitung der Aufgaben ausreicht (u. a. Amelang & Zielinski, 2002; Jonkisz et al., 2012).

Um zu gewährleisten, dass die Ergebnisse dieses Leistungstests nicht durch die Beurteilenden beeinflusst sind, wurde ein **objektiver Zugang** zur Erfassung der Modellkompetenz gewählt (Jonkisz et al., 2012; Kunter & Klusmann, 2010). Dieser ist dadurch charakterisiert, dass eine Messung **von außen** anhand ex-

⁴ Zum Zeitpunkt der Befragung (2009 bis 2011) gab es noch die Schulform **Realschule** in Berlin. Im Schuljahr 2010/2011 wurden in Berlin Haupt-, Real- und Gesamtschulen zu integrierten Sekundarschulen zusammengesfasst.

terner Kriterien erfolgt. Diese Kriterien (Validität, Reliabilität und Objektivität) werden in allen Phasen der Testkonstruktion entsprechend berücksichtigt. Unter **Objektivität** wird die Unabhängigkeit eines Testergebnisses von anderen Einflüssen verstanden, welche sich in die Interpretations-, die Auswertungs- und die Durchführungsobjektivität aufteilt. **Reliabilität** ist das Maß für die Zuverlässigkeit und Genauigkeit einer Messung (Rost, 2004). Für qualitative Untersuchungen werden beispielsweise Interkoder-Reliabilitäten berechnet, die Aussagen über die Übereinstimmung zwischen zwei Kodierenden ermöglichen (Gläser-Zikuda, 2005; Mayring, 2010). Das Kriterium der **Validität** wird in dieser Arbeit nicht als eine Eigenschaft von Tests verstanden, die sich in Inhalts-, Kriteriums- oder Konstruktvalidität aufgeteilt. Vielmehr orientiert sich die Untersuchung an dem aktuellen Ansatz, Validität als die Gültigkeit von Testwertinterpretationen zu definieren (vgl. Frey, 2014; Hartig et al., 2012; Leuders, 2014; Schmiemann & Lücken, 2014). Das bedeutet, dass zunächst entschieden werden muss, welche Interpretationen für den Test relevant sind. Anschließend werden diese Interpretationen in den einzelnen Phasen der Testkonstruktion mit theoretischen Argumenten und empirischen Befunden gestützt (vgl. Kane, 2001, 2013). Mit diesem integrativen Validitätskonzept werden nun nicht mehr unterschiedliche Eigenschaften der Validität getrennt voneinander betrachtet, sondern verschiedene Eigenschaften werden auf intendierte Interpretationsschlüsse hin fokussiert (vgl. Hartig et al., 2012; Leuders, 2014; Schmiemann & Lücken, 2014).

Als Messinstrument werden im Rahmen der vorliegenden Forschungsarbeit **Aufgaben im offenen Antwortformat** eingesetzt. Es werden also keine Antwortalternativen vorgegeben, die getesteten Personen müssen daher eigenständig eine Antwort formulieren. Wenn eine Antwort selbständig zu formulieren ist, ist sie weniger durch Vorgaben beeinflusst und eine zufälligerweise richtige Antwort der Testperson ist weniger wahrscheinlich (Jonkisz et al., 2012; Rost, 2004). Der Einsatz offener Antwortformate hat sich bereits bei der Überprüfung von anderen Kompetenzmodellen (vgl. Eggert & Bögeholz, 2010; Grube, 2011) und bei der Erfassung von konkreten und abstrakten Vorstellungen von Modellen bewährt (vgl. Crawford & Cullin, 2005; Sins et al., 2009; Van Driel & Verloop, 1999). Daher eignet sich dieses Aufgabenformat dazu, verschiedene konkrete und abstrakte Perspektiven von Schülerinnen und Schülern auf Modelle und auf den Modellbildungsprozess zu erfassen. Diese Perspektiven der

Schülerinnen und Schülern können für eine empirische Überprüfung des Kompetenzmodells herangezogen werden. Darüber hinaus dienen die verschiedenen Perspektiven dazu, einzelne Niveaus genauer zu beschreiben und auszu-differenzieren.

Neben der Erfassung von Perspektiven eignet sich das offene Antwortformat auch dafür, aus den Aussagen der Lernenden Besonderheiten zu identifizieren, mit denen sich die Ergebnisse der Analysen erklären lassen. Interviews⁵ lassen zwar einen detaillierteren Einblick in die Perspektiven von Schülerinnen und Schülern zu als Aufgaben im offenen Format, sie sind in der Auswertung jedoch aufwändiger. Für das Ziel dieser Arbeit, das Kompetenzmodell mit einer großen Anzahl von Lernenden aus verschiedenen Jahrgangsstufen zu überprüfen, sind Aufgaben im offenen Antwortformat Interviews vorzuziehen (vgl. Rost, 2004).

4.1.3 Aufgabenformat und -auswertung

Die Wahl des Aufgabenformats hat immer Konsequenzen für die Durchführung, die Auswertung, die Interpretation und die Ökonomie eines Tests (Lienert & Raatz, 1998). Es lassen sich freie (oder offene) und gebundene Aufgabenformate unterscheiden (vgl. Rost, 2004). Bei freien Aufgabenformaten, also solchen, bei dem die zu testende Person selbst eine Antwort formuliert, entsteht der zu testenden Person ein hoher Bearbeitungsaufwand und den Auswertenden ein hoher Auswertungsaufwand. Gebundene Aufgabenformate dagegen sind durch eine Auswahl aus Antwortalternativen gekennzeichnet und in der Durchführung und Auswertung einfach, ökonomisch und objektiv (Hammann & Jördens, 2014; Jonkisz et al., 2012; Rost, 2004). Außerdem kann der Rückschluss von der Bearbeitung einer Aufgabe im freien Format auf die Kompetenz einer Person (*Testwertinterpretation*, vgl. Hartig et al., 2012) eingeschränkt sein, beispielsweise wenn eine Testperson Formulierungsschwierigkeiten hat (u. a. Jonkisz et al., 2012). Letzteres ist für die vorliegende Forschungsarbeit nicht relevant, weil hier nicht einzelne Lernende diagnostiziert werden sollen, sondern allgemein die Erfassung der Perspektiven von Lernen-

⁵ Um einen detaillierten Einblick in die Vorstellungen von Schülerinnen und Schülern zu gewinnen, führten Trier und Upmeyer zu Belzen (2009) in einem parallelen Projekt Interviews mit Schülerinnen und Schülern durch.

den und Aussagen auf Populationsebene angestrebt werden. Um dennoch einen legitimen Rückschluss auf die Kompetenz gewährleisten zu können, wurde die Testlänge, also die Anzahl der Aufgaben pro Testheft, auf drei bis fünf Aufgaben begrenzt (Grünkorn et al., 2011; Grünkorn & Krüger, 2012). Diese geringe Anzahl von Aufgaben wurde einerseits gewählt, um den Bearbeitungsaufwand auf Seiten der zu testenden Personen gering zu halten, und andererseits sollte dies es den Lernenden ermöglichen, ihre Antworten ausführlich zu begründen, damit die Auswertenden diese Aussagen eindeutig zuordnen konnten. Auf diese Weise kann die Auswertungs- und Interpretationsobjektivität erhöht werden (vgl. Hartig et al., 2012; Kauertz et al., 2010).

Weiterhin sind für eine objektive Auswertung einzelne **Auswertungsschritte** festzulegen. Als Methode dafür ist die qualitative Inhaltsanalyse (Mayring, 2010) geeignet, die im Rahmen dieser Forschungsarbeit computergestützt mit der Software MAXQDA (Version 10) durchgeführt wurde. Um der Mehrdeutigkeit von Aussagen zu begegnen und die Objektivität der Auswertung zu gewährleisten, wurde für jede Teilkompetenz ein Kodierleitfaden (Anhang 1) konstruiert (vgl. Hammann & Jördens, 2014; Jonkisz et al., 2012; Mayring, 2010; Rost, 2004). Dadurch können die Kodierungen sowohl verglichen als auch überprüft werden. In die Entwicklung und Überprüfung dieser Kodierleitfäden flossen die Daten aus den Teilstudien 1 (Erhebung 2009) und 2 (Erhebung 2010/2011) ein (Grünkorn et al., 2011; Grünkorn & Krüger, 2012). In Abbildung 6 sind die Auswertungsschritte skizziert, auf die im Folgenden eingegangen wird.

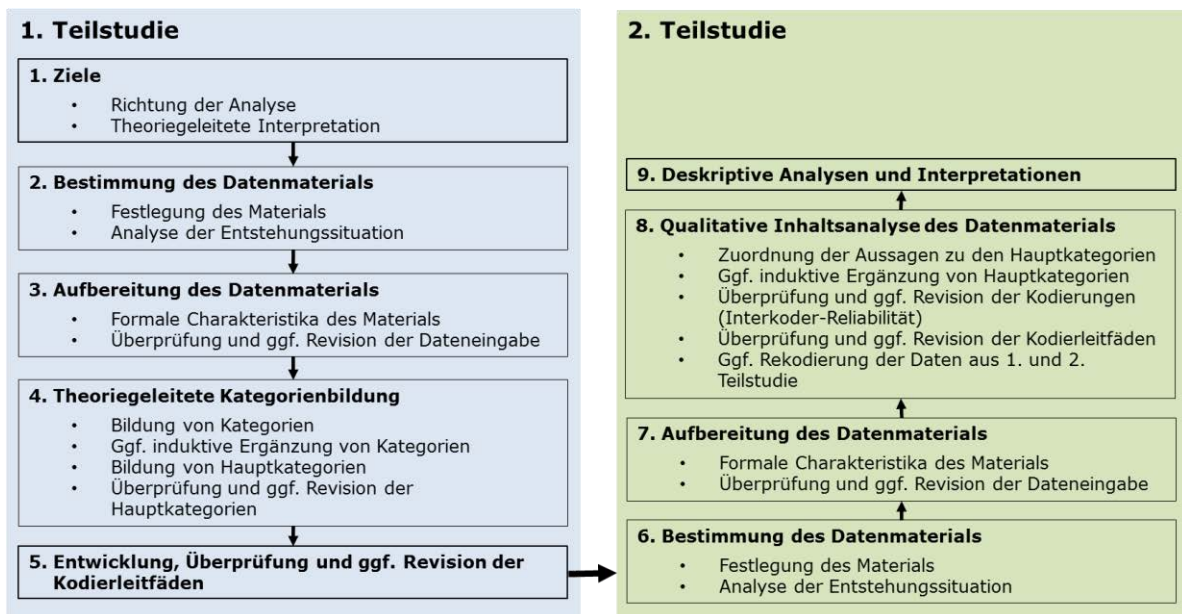


Abb. 6: Auswertungsschritte der Teilstudien 1 und 2 zur Operationalisierung (vgl. Gläser-Zikuda, 2005; Gropengießer, 2005; Mayring, 2010)

Bevor das Datenmaterial aufbereitet und ausgewertet wird, sind die *Richtung der Analyse* und die *theoriegeleitete Interpretation* zu bestimmen und zu beschreiben (Abb. 6; Mayring, 2010). Ziel der Teilstudien 1 und 2 war es, Aufgaben zu entwickeln, die valide als Indikatoren der jeweiligen Teilkompetenz des Kompetenzmodells interpretiert werden können (vgl. Kane, 2001). Somit lag der Fokus darauf, zu überprüfen, inwiefern die Aufgabenstämme den Lernenden plausibel und verständlich waren und inwiefern die Kontexte, Modelle und Stimuli Antworten auslösten, die sich auf die jeweilige Teilkompetenz und deren Spektrum bezogen (Qualität der Aufgaben). Außerdem wurden in diesen Studien bereits verschiedene Perspektiven auf Modelle und auf den Modellbildungsprozess identifiziert. Auf dieser Basis war es möglich, Kodierleitfäden zu entwickeln und schließlich das Hauptziel der Forschungsarbeit, die empirische Überprüfung des Kompetenzmodells der Modellkompetenz, zu erreichen. Um die inhaltlichen Aspekte im Kompetenzmodell mit denen der Lernenden zu vergleichen, müssen die entwickelten Aufgaben die Lernenden dazu anregen, über Modelle und über den Modellbildungsprozess nachzudenken und ihre Perspektiven zu dokumentieren. Somit gab es zwei Analyseschwerpunkte: Zum einen wurde analysiert, inwiefern sich auf der Grundlage der von den Lernenden verfassten Antworten Aussagen über die Qualität der Aufgaben treffen ließen, und zum anderen wurde analysiert, inwiefern aus den Antworten der Lernenden

verschiedene Perspektiven auf Modelle und auf den Modellbildungsprozess ersichtlich waren (*Richtung der Analyse*, Mayring, 2010). Zusammen mit dieser Richtung der Analyse orientierte sich die Interpretation der Daten an der Beschreibung und Definition der jeweiligen Teilkompetenz (vgl. Upmeyer zu Belzen & Krüger, 2010).

Als zweiter Schritt erfolgte die *Bestimmung des Datenmaterials* (Abb. 6; Mayring, 2010) der Teilstudie 1. Hier handelte es sich um schriftliche Aussagen von $N = 510$ Schülerinnen und Schülern (Realschule, 7. Jgst. und Gymnasium, 10. Jgst.) zu Aufgaben einer spezifischen Teilkompetenz des Kompetenzmodells. Bei der Stichprobenauswahl stand die Breite des Antwortspektrums im Vordergrund: Alle Aussagen aus dieser Stichprobe wurden aufbereitet, es wurde also keine Auswahl getroffen. Die Teilnahme an der Studie war freiwillig und anonym. Nach einem Informationsschreiben und einem Informationsgespräch mit den Schulleitungen wurden vor der Testung Einverständniserklärungen der Eltern und der Schulleitungen eingeholt. Die Testungen fanden innerhalb der regulären Unterrichtszeit statt, benötigten etwa einer Schulstunde und wurden von der Autorin selbst oder von Testleiterinnen und Testleitern durchgeführt. Um die Durchführung zu objektivieren, wurden die Testleiterinnen und Testleiter in einem Gespräch über den Ablauf informiert. Zudem führten sie die Testung nach einem Testmanual (Anhang 2) durch und führten ein Protokoll darüber (z. B. Störungen im Ablauf, Fragen der Lernenden, gegebene Antworten; vgl. Jonkisz et al., 2012; Neuhaus & Braun, 2007; Rost, 2004).

Im dritten Schritt erfolgte die *Aufbereitung des Datenmaterials* (Abb. 6; Mayring, 2010): Die Aussagen der Lernenden, die bereits schriftlich vorlagen, wurden manuell in die Software MAXQDA übertragen. In diesem Auswertungsschritt mussten nicht nur die vollständigen Äußerungen der einzelnen Lernenden digitalisiert, sondern auch alle von ihnen gesetzten Anmerkungen, Pfeile und Zeichnungen verschriftlicht und übertragen werden. Unklarheiten hinsichtlich der Lesbarkeit des Textes wurden markiert und mit einer weiteren Person besprochen. Rechtschreibfehler wurden korrigiert. Aussagen, die schwer zu verstehen waren, wurden so korrigiert, dass sie an Klarheit und Deutlichkeit gewannen, ohne dabei allerdings den Sinn und den Stil der Aussage zu verändern. Notwendige Änderungen bzw. Ergänzungen wurden im Transkript mit eckigen Klammern gekennzeichnet (vgl. Krüger & Riemeier, 2014; Mayring, 2010). Nach Abschluss der Dateneingabe nach diesen Regeln wurden die Tran-

skripte vollständig von einer Person, die mit der Dateneingabe und den entsprechenden Regeln vertraut war, überprüft und gegebenenfalls verändert.

Der vierte Schritt war die **theoriegeleitete Kategorienbildung** (Abb. 6; vgl. Gläser-Zikuda, 2005; Hammann & Jördens, 2014; Krüger & Riemeier, 2014; Mayring, 2010). Diese erfolgte unter Berücksichtigung

- der Definitionen der einzelnen Teilkompetenzen (Upmeier zu Belzen & Krüger, 2010),
- der Beschreibungen der einzelnen Niveaus (Upmeier zu Belzen & Krüger, 2010) sowie
- der beiden theoriegeleitet definierten Perspektiven auf Modelle (Gilbert, 1991; Mahr, 2009): Modelle als Medien (Modellobjektebenen- und Herstellungsperspektive) und Modelle als Methode (Anwendungsperspektive).

Konnten Aussagen der Lernenden keinem Niveau der jeweiligen Teilkompetenz zugeordnet werden, wurden induktiv Kategorien entwickelt und ergänzt (vgl. Hammann & Jördens, 2014). Bei der Auswertung spielten weder die fachliche Richtigkeit noch die adäquate Planung einer Untersuchung eine Rolle: Solche Aussagen flossen unkorrigiert in die Auswertung ein und wurden nicht selektiert. Die Analyse begann mit der Gruppierung der Aussagen derjenigen Lernenden, die entweder die Aufgaben unbearbeitet ließen (Hauptkategorie **keine Antwort**) oder deren Antworten nicht mit der Frageintention korrespondierten (Hauptkategorie **Sonstiges**). Die Eingruppierung in die Hauptkategorie **Sonstiges** ergab hilfreiche Informationen hinsichtlich der Verständlichkeit der Aufgaben, und je nach Teilkompetenz und Aufgabe wurden hierunter unterschiedliche Kategorien subsumiert. Für die Aufgabe **Enzymatische Prozesse im Mund** der Teilkompetenz **Zweck von Modellen** wurden beispielsweise folgende Kategorien unter die Hauptkategorie **Sonstiges** subsumiert:

- Kategorie **Keine Idee**: „Weiß ich nicht, ich habe keine Idee.“ (Fa381)
- Kategorie **Keine Korrespondenz zur Frageintention**: „Die Biologinnen haben es so erschaffen.“ (Fa199)
- Kategorie **Hinweise zum Verständnis**: „Ohne Bild versteht man die Frage gar nicht.“ (Fa435)

- Kategorie **Scherz-Antwort**: „Dieses Modell sieht aus wie eine grüne Kette mit lila Diamanten. Das Modell hat den Zweck, den Hals der Frau zu verschönern.“ (Fa354)

Anschließend erfolgte mit Hilfe der den Frageintentionen entsprechenden Antworten der Lernenden die theoriegeleitete Bildung von Kategorien. Solche Kategorien orientieren sich an konkreten Antworten pro Aufgabe und weisen einen geringen Abstraktionsgrad auf (vgl. Gläser-Zikuda, 2005). Für die Kategorienbildung wurden minimal ganze Sätze (Kodiereinheit) und maximal ganze Antworten von Schülerinnen und Schülern (Kontexteinheit) herangezogen (vgl. Mayring, 2010). Jede gebildete Kategorie fasst Antworten zusammen, die eine ähnliche Perspektive hinsichtlich einer spezifischen Teilkompetenz und eines spezifischen Niveaus beschreiben (vgl. Krüger & Riemeier, 2014). Für das Niveau I der Aufgabe **Biomembran** der Teilkompetenz **Alternative Modelle** wurden beispielsweise folgende Kategorien identifiziert (Grünkorn et al., 2011, 2014):

- Kategorie **Verschiedene Darstellungsweisen**: „Modell A ist zum Beispiel nicht 3D [...]. Modell B und C sind 3D und unterschiedlich groß [...].“ (Fb44)
- Kategorie **Verschiedene Modellmerkmale**: „Modell B und C werden unterschieden durch ihre Beschaffenheit. Während Modell B eher steif ist, ist Modell C sehr flexibel.“ (Fb64)
- Kategorie **Unterschiedliche Konstruktionsmöglichkeiten**: „Bei Modell A sind die verschiedenen Bausteine gut zu erkennen. [...] Für mich sind die Stäbchen auf Modell C dünner [als] bei B.“ (Fb311)

Aus den gesamten Kategorien eines Niveaus der jeweiligen Teilkompetenz wurden theoriegeleitet Hauptkategorien generiert. Die Hauptkategorien weisen ein höheres Abstraktionsniveau auf als die Einzelkategorien und verallgemeinern die konkreten Kategorien der einzelnen Aufgaben (vgl. Gläser-Zikuda, 2005). Beim genannten Beispiel der Aufgabe **Biomembran** der Teilkompetenz **Alternative Modelle** etwa wurden die drei aufgeführten Kategorien zur Hauptkategorie **Unterschiedliche Modellobjekteigenschaften** zusammengefasst. Um die gebildeten Hauptkategorien zu überprüfen und gegebenenfalls zu revidieren, wurden mit Teilen des Materials in parallelen Studien zusätzlich Kategorien und Hauptkategorien für jede Teilkompetenz generiert (**Eigenschaften**

von Modellen, Ciechorski, 2009; *Alternative Modelle*, Koslowski, 2009; *Zweck von Modellen*, Zeybak, 2010; *Testen und Ändern von Modellen*, C. Windisch, persönliche Mitteilung, 03.03.2010). Die Kategorisierungen unterschieden sich vor allem hinsichtlich des gewählten Abstraktionsgrads und der Zusammenfassung der Kategorien zu Hauptkategorien. Daher wurden nur die einzelnen Kategorien und nicht die Hauptkategorien für die Diskussionen herangezogen. Die Möglichkeit, die eigenen Kategorien mit denen aus parallelen Arbeiten abzugleichen, stützt darüber hinaus die Legitimität der Formulierung zusätzlicher Kategorien zu den Teilkompetenzen des Kompetenzmodells (vgl. Grünkorn & Krüger, 2012; Grünkorn et al., 2011), die unabhängig voneinander identifiziert wurden. Zudem konnten die Hauptkategorien präzisiert werden, wodurch sie an Trennschärfe gewannen. Über diesen Abgleich hinaus wurden die Hauptkategorien mit acht Biologiedidaktikerinnen und Biologiedidaktikern, die Expertinnen und Experten auf dem Gebiet der Modellkompetenz sind, diskutiert, und die Hauptkategorien wurden gegebenenfalls revidiert (*formative Reliabilitätsprüfung*, vgl. Gläser-Zikuda, 2005). Die Diskussion bezog sich auf vier Fragestellungen: (1) Inwiefern spiegeln die Hauptkategorien das Niveau der jeweiligen Teilkompetenz wider? (2) Inwiefern spiegeln die gezeigten Ankerbeispiele die Perspektive der Hauptkategorie wider? (3) Inwiefern sind die Hauptkategorien präzise genug definiert? (4) Inwieweit grenzen sie sich damit von den anderen Hauptkategorien ab? Für beide Qualitätsschritte wurde auf eine Berechnung der Interkoder-Reliabilitäten verzichtet, da in der ersten Teilstudie zunächst Kategorien identifiziert und Hauptkategorien generiert werden sollten. In einem fünften und abschließenden Schritt der Teilstudie 1 wurden für die einzelnen Teilkompetenzen *Kodierleitfäden* entwickelt (Abb. 6). Diese enthalten Informationen zum theoretischen Rahmen und zu den Kodierregeln sowie ein Kodierschema, das für alle Hauptkategorien eine Kurzbezeichnung enthält, diese definiert und durch beispielhafte Aussagen von Schülerinnen und Schülern veranschaulicht (Anhang 1).⁶ Diese Kodierleitfäden wurden ebenfalls mit den zuvor genannten Expertinnen und Experten auf dem Gebiet der Modellkompetenz diskutiert. Die Diskussion bezog sich auf die Verständlichkeit der Informationen und der Regeln, die adäquate Wahl von Ankerbeispielen und die genaue Abgrenzung der Beschreibungen der einzelnen Hauptkategorien.

⁶ In den Kodierleitfäden wird keine Unterscheidung in Kategorie und Hauptkategorie vorgenommen. Wird von einem Kode gesprochen, ist damit stets die Hauptkategorie gemeint.

Um die entwickelten Kodierleitfäden zu überprüfen, gegebenenfalls weitere Hauptkategorien induktiv zu ergänzen und die überarbeiteten, aber auch zusätzlich entwickelte Aufgaben zu evaluieren, wurde eine zweite Teilstudie durchgeführt. Unter Berücksichtigung der Theorie und der Fragestellungen (Abb. 6) erfolgte zunächst die **Bestimmung des Datenmaterials** (Mayring, 2010) der Teilstudie 2. Dieses entsprach mit Ausnahme der Anzahl der getesteten Personen dem der Teilstudie 1. Es handelte sich um schriftliche Aussagen von $N = 721$ Schülerinnen und Schülern (Realschule, 7. bis 10. Jgst.; Gymnasium, 8. bis 10. Jgst.) zu Aufgaben im offenen Antwortformat.

Nach der **Aufbereitung des Datenmaterials** (vgl. Teilstudie 1) kam es in einem achten Schritt zur **qualitativen Inhaltsanalyse des Datenmaterials** (vgl. Gläser-Zikuda, 2005). Auf der Basis der Kodierleitfäden wurden die Daten von mindestens drei Kodierenden (einschließlich der Autorin) ausgewertet, die mit dem Kompetenzmodell der Modellkompetenz, den Kodierregeln und dem Kodierschema vertraut waren. Die Kodierenden hatten die Aufgabe, die Aussagen mit Hilfe des Kodierleitfadens den Hauptkategorien zuzuordnen, gegebenenfalls induktiv neue Kategorien zu ergänzen und Schwierigkeiten mit dem Kodierleitfaden zu benennen. Anschließend mussten die Kodierleitfäden revidiert und gegebenenfalls die Daten aus der ersten und zweiten Teilstudie rekodiert werden. Unterschiede bei den Zuordnungen bzw. neue Kategorien wurden so lange diskutiert, bis ein Konsens erreicht wurde (**konsensuelle Validierung**, Flick, 2002). An dieser Stelle wurde die Berechnung von Cohens Kappa (κ) vorgenommen. Dieser Wert bestimmt das Ausmaß der Übereinstimmung zwischen zwei Kodierenden (Bortz & Döring, 2006). Eine Übereinstimmung ist dann erreicht, wenn beide Kodierende dieselbe Textpassage innerhalb einer Aussage derselben Hauptkategorie zuordnen. Als Unterschiede werden Kodierungen gewertet, bei denen dieselbe Textpassage unterschiedlichen Hauptkategorien zugeordnet wird.

Mit den überprüften Kodierungen konnte die Vorbereitung (Erstellung einer Excel-Matrix) für **deskriptive Analysen und Interpretationen** der Daten aus den Teilstudien 1 und 2 getroffen werden (vgl. Gläser-Zikuda, 2005). Diese Matrix enthält pro Teilkompetenz, Aufgabe und Person Informationen darüber, welche Hauptkategorien von den Lernenden genannt wurden sowie welche Niveaus insgesamt und welches höchste Niveau erreicht wurden. Anschließend wurden Häufigkeiten mit SPSS berechnet.

4.1.4 Aufgabenentwicklung

Um die Aufgabenentwicklung durch eine systematische Vorgehensweise zu objektivieren, wurde die Entwicklung der Aufgaben pro Teilkompetenz standardisiert durchgeführt und in einem Leitfaden dokumentiert (Anhang 3). Dieser Leitfaden wurde im Laufe der Aufgabenentwicklung optimiert und ergänzt. Er beschreibt pro Teilkompetenz den theoretischen Bezug und gibt Informationen zur Struktur des Aufgabenstamms, zum standardisierten Stimulus, zu Konstruktionshinweisen und zu geeigneten Kontexten. Tabelle 3 zeigt exemplarisch für die Teilkompetenz *Eigenschaften von Modellen* eine Konstruktionsanleitung, die über die vorliegende Untersuchung hinaus in anschließenden Studien für die Konstruktion weiterer Items genutzt werden kann. Das Beispiel für eine Konstruktionsanleitung (Tab. 3) veranschaulicht zugleich einige Charakteristika der entwickelten Aufgaben, auf die im Folgenden näher eingegangen wird:

- *Repräsentation des Spektrums* der jeweiligen Teilkompetenz
- *Standardisierter Aufbau* der Aufgaben
- *Kontextauswahl*: Berücksichtigung weiterer schwierigkeiterzeugender Merkmale, wie *Aufgabenformat*, *formale Merkmale* und *Fachwissen*.

Das erste Charakteristikum, die *Repräsentation des Spektrums* der jeweiligen Teilkompetenz durch die Aufgaben, ist eine notwendige Voraussetzung für die valide Interpretation der Antworten der Lernenden als Indikatoren der jeweiligen Teilkompetenz (vgl. Kane, 2001). Dies ist nur dann möglich, wenn der gewählte Aufgabenkontext, das Modell und der Stimulus Antworten auf allen drei Niveaus der jeweiligen Teilkompetenz ermöglichen und damit beide Perspektiven auf Modelle (*Modell als Medium* und *Modell als Methode*, Gilbert, 1991; Mahr, 2009) miteinander verbinden. So muss beispielsweise eine Aufgabe mit Bezug auf die Teilkompetenz *Eigenschaften von Modellen* solche Antworten von Lernenden zulassen, die Modelle als Kopien (Niveau I), als idealisierte Repräsentationen (Niveau II) oder als theoretische Rekonstruktionen (Niveau III) beschreiben (vgl. Tab. 3).

Tab. 3: Anleitung zur Aufgabenkonstruktion für die Teilkompetenz *Eigenschaften von Modellen*

Eigenschaften von Modellen (Niveau I bis III)	
Theoriebezug	<ul style="list-style-type: none"> - Schilderung der Beziehung zwischen Modell und Original - Perspektiven auf Modelle als Medien und als Methode: Modelle als Kopien (Niveau I), idealisierte Repräsentationen (Niveau II) oder theoretische Rekonstruktionen (Niveau III)
Struktur des Aufgabensamms	<ol style="list-style-type: none"> 1. Hinweis geben, dass das Original nicht direkt beobachtbar und nur mit Hilfsmitteln erkennbar ist (z. B. Röntgengerät, Mikroskop). 2. Daten beschreiben, die die Grundlage für die Konstruktion des Modells sind. 3. Grafische Darstellung des Prozesses der Entstehung des Modells: Bilder vom Original bzw. von den Untersuchungsergebnissen klar als Datengrundlage kennzeichnen. 4. Abbildung vom Modell zeigen.
Standardisierter Stimulus	<ul style="list-style-type: none"> - Beschreibe, inwieweit das Modell des [Originals] so aussieht wie das [Original]. Begründe deine Antwort!
Konstruktionshinweise	<ul style="list-style-type: none"> - Die gezeigten Originale stellen nicht selbst Modelle dar: Beispielsweise anstatt eines zusammengesetzten Knochenfundes gefundene Knochenreste oder Fossilien zeigen. - Nicht zu komplexe oder abstrakte Abbildungen des Originals wählen. Röntgenbilder beispielsweise sind weniger gut geeignet. - Wenn möglich nur ein Modell und passend dazu die jeweiligen Abbildungen des Originals zeigen, damit die Lernenden Modell und Original vergleichen. Werden mehrere Modelle gezeigt, neigen Lernende dazu, die Modelle miteinander statt das Original mit einem Modell zu vergleichen. - Original und Modell unterscheiden sich sichtbar voneinander bzw. grenzen sich ab, damit ein Interpretationsspielraum – auch für Niveau III – möglich ist.
Geeignete Kontexte	<ul style="list-style-type: none"> - Modelle, deren Originale nicht direkt beobachtbar sind, zum Beispiel Mikrokosmos, Evolution

Aufgaben zu konstruieren, die sich auf eine spezifische Teilkompetenz und nur ein Niveau beziehen, wie dies bei gebundenen Aufgabenformaten häufig geschieht (z. B. MC-Items, Terzer, 2013), ist bei freien Antwortformaten schwierig. Modelle können immer als Medium und als Methode betrachtet werden (Gilbert, 1991; Mahr, 2009). Infolgedessen können die Lernenden bei freien Aufgabenformaten zur Modellkompetenz auch beide Perspektiven formulieren. Somit ist eine trennscharfe Erfassung pro Niveau ohne die Vorgabe von Antwortalternativen nicht möglich. Wie bei dieser Struktur offener Formate bezie-

hen sich auch die von Krell und Krüger (2010; Krell, Czeskleba et al., 2012) entwickelten FC-Items auf alle drei Niveaus der jeweiligen Teilkompetenz.

Als zweites Charakteristikum ist der **Aufbau der Aufgaben** standardisiert, das heißt, der Aufbau des Aufgabenstamms und der standardisierte Stimulus sind analog (vgl. Tab. 3; Grünkorn, Upmeier zu Belzen et al., 2014; Grünkorn et al., 2011). Dies ist deshalb von Bedeutung, weil anzunehmen ist, dass eine unterschiedliche Gestaltung und verschiedene Stimuli innerhalb einer Teilkompetenz Einfluss auf die Lösungswahrscheinlichkeit haben (vgl. Jonkisz et al. 2012; Rost, 2004; Terzer et al., 2013). Aus diesem Grund wurden in einem Leitfaden Hinweise auf eine standardisierte Entwicklung von Aufgaben gegeben (Anhang 3; vgl. Tab. 3; vgl. Rost, 2004): Jede Aufgabe besteht aus einem Aufgabenstamm, der wesentliche fachliche Informationen gibt und Abbildungen des Modells und in manchen Fällen auch des Originals zeigt. Der Aufgabenstamm schließt mit einem standardisierten Stimulus ab, der durch einen Rahmen vom restlichen Text abgehoben ist. Das Antwortformat ist offen, weshalb Platz für eine Antwort freigehalten ist (Beispielaufgaben: vgl. Grünkorn, Upmeier zu Belzen et al., 2014; Grünkorn et al., 2011). Die Aufgabenstellung pro Teilkompetenz ist standardisiert, um Unterschiede im Antwortverhalten durch verschiedene Formulierungen auszuschließen (vgl. Terzer et al., 2013). Um Stimuli anzubieten, die von den Lernenden verstanden werden und Antworten auf allen drei Niveaus erlauben, wurden die Fragen aus den Interviewstudien von Grosslight et al. (1991) sowie Trier und Upmeier zu Belzen (2009) als primäre Quellen zur Entwicklung der Aufgabenstellungen herangezogen. Die Offenheit der entwickelten Stimuli (vgl. Grünkorn & Krüger, 2012; Grünkorn et al., 2011) ermöglicht es, das gesamte Spektrum einer Teilkompetenz abzudecken.

Das dritte Charakteristikum betrifft die **Kontextauswahl**. Dabei bezieht sich der Begriff **Kontext** auf Inhalte, die im Aufgabenstamm mit Modellen und Fachinformationen beschrieben werden. In der Kompetenzdefinition von Klieme und Leutner (2006) ist die Kontextspezifität verankert, so dass sie bei der Operationalisierung beachtet werden muss. Darüber hinaus ist es zweckmäßig, unterschiedliche biologische Kontexte (u. a. Evolution, Ökologie, Zoologie, Humanbiologie, Zellbiologie) zu wählen, weil Modelle in der Biologie für verschiedene Themenbereiche relevant sind. Durch unterschiedliche Kontexte kann das Konstrukt angemessen repräsentiert werden, was wiederum eine valide Verallge-

meinerung der Testergebnisse von den Testinhalten auf das Konstrukt Modellkompetenz ermöglicht (vgl. Hartig et al., 2012). Bei der Auswahl der Kontexte spielte eine systematische Variation der biologischen Kontexte keine große Rolle, vielmehr wurde darauf geachtet, dass die Kontexte Antworten auf den drei Niveaus der jeweiligen Teilkompetenz erlaubten und dass sie leicht erklärbar und damit für die Zielgruppe (7. bis 10. Jahrgangsstufe Realschule und Gymnasium) verständlich waren (vgl. Kane, 2001). So geht es beispielsweise bei der Teilkompetenz **Eigenschaften von Modellen** darum, ob Modelle als Kopien (Niveau I), als idealisierte Repräsentationen (Niveau II) oder als theoretische Rekonstruktionen (Niveau III) wahrgenommen werden. Um plausible Antworten auf Niveau I, II und III zu ermöglichen, sind Kontexte geeignet (z. B. Mikrokosmos, Evolution), deren Originale nicht (mehr) direkt beobachtbar sind, so dass den Lernenden der hypothetische Charakter von Modellen deutlich wird. Die gewählten Aufgabenkontexte beschreiben möglichst konkrete biologische Situationen, in denen der Einsatz von Modellen realistisch nachvollziehbar ist. Dadurch, dass die Situationen möglichst alltagsnah gewählt sind, lässt sich die Testmotivation der Lernenden steigern (u. a. Hammann, 2006).

Als abschließendes Charakteristikum sind weitere schwierigkeiterzeugende Merkmale, wie das Aufgabenformat (vgl. Hartig & Klieme, 2006; Martinez, 1991), formale Merkmale (vgl. Jonkisz et al., 2012; Kauertz et al., 2010; Prenzel et al., 2002; Rost, 2004) und das Fachwissen (vgl. Kauertz et al., 2010; Prenzel, Häußler, Rost & Senkbeil, 2002) minimiert oder konstant gehalten. Eine vollständige Kontrolle dieser Faktoren ist allerdings nicht möglich (Prenzel et al., 2002). So kann bereits das **Aufgabenformat** Einfluss auf das Antwortverhalten haben (vgl. Hartig & Klieme, 2006; Martinez, 1991). Untersuchungen haben gezeigt, dass die Offenheit des Aufgabenformats die Schwierigkeit der Aufgabe beeinflusst (Martinez, 1991). Die Ergebnisse aus Studien zur Aufgabenschwierigkeit hierzu sind uneinheitlich, so dass sich nicht grundsätzlich sagen lässt, freie Aufgabenformate seien schwieriger als gebundene (Bennett & Ward, 1993; Traub, 1993). Um diesen Faktor konstant zu halten, wurden im Rahmen dieser Forschungsarbeit Aufgaben im offenen Antwortformat eingesetzt und auch nur diese bei der Auswertung berücksichtigt.

Weiterhin wurden bei der Aufgabenentwicklung solche **formalen Merkmale** wie die Textlänge und die Sprache beachtet (vgl. Jonkisz et al., 2012; Kauertz et

al., 2010; Prenzel et al., 2002; Rost, 2004). Um den Einfluss der Lesefähigkeit zu minimieren, sind die Fachinformationen im Aufgabenstamm kurz und einfach formuliert (vgl. Jonkisz et al., 2012; Lienert & Raatz, 1998; Neuhaus & Braun, 2007). Zudem wurden Fremdwörter, doppelte Verneinungen sowie mehrdeutige Begriffe vermieden, um die Sprache der Lernenden zu berücksichtigen (Neuhaus & Braun, 2007).

Modellkompetenz ist im Bereich der Erkenntnisgewinnung angesiedelt. Die entwickelten Aufgaben sollten daher kein **Fachwissen** erfassen. Damit die Lernenden gegebenenfalls auch bei Unkenntnis des Themas die notwendige Wissensbasis zur Bearbeitung der Aufgaben zur Verfügung haben, sind nur unbedingt notwendige Fachinformationen im Aufgabenstamm beschrieben (Kauertz et al., 2010). Um die Verständlichkeit zu unterstützen, werden Abbildungen der Modelle und zum Teil auch der Originale gezeigt. Darüber hinaus wurden auch diejenigen Antworten ausgewertet, die mit Blick auf den biologischen Inhalt fachlich nicht korrekt waren. Auch dadurch lässt sich der Einfluss des Faktors Fachwissen minimieren (vgl. Prenzel et al., 2002).

4.1.5 Aufgabenüberprüfung und -selektion

Nach der Aufgabenentwicklung ist zu überprüfen, inwiefern die Aussagen der Lernenden zu den entwickelten Aufgaben als Indikatoren der jeweiligen Teilkompetenz des Kompetenzmodells interpretiert werden können (vgl. Kane 2001). Dies ist deshalb wichtig, weil nur dann valide auf die Kompetenz von Lernenden geschlossen werden kann (vgl. Hartig et al., 2012). Für die Aufgabenüberprüfung und -selektion wurden im Vorhinein Kriterien festgelegt, anhand derer die Qualität der Aufgaben beurteilt wurde:

- Plausibilität und Verständlichkeit des Aufgabenstamms – mit Blick auf fachliche Informationen, Modell und Stimulus – für die gewählte Zielgruppe
- Bezug der Aussagen der Lernenden zur jeweiligen Teilkompetenz (**ein-deutige Modelleinordnung**, Kauertz et al., 2010)
- Repräsentation der drei Niveaus der jeweiligen Teilkompetenz in den Aussagen der Lernenden (**Modellsättigung**, Kauertz et al., 2010).

Im Hinblick auf diese Kriterien sind die Aussagen der Lernenden zu den Aufgaben als Indikatoren von Modellkompetenz anzusehen, wenn die Aufgaben den Lernenden der Zielgruppe (Realschule und Gymnasium, 7.–10. Jgst.)

- ein plausibles Modell bieten,
- die dazu gebotenen Fachinformationen verständlich sind und
- der Kontext, das Modell und der Stimulus Antworten der Lernenden erlauben, die sich auf die jeweilige Teilkompetenz und deren Spektrum beziehen (vgl. Grünkorn & Krüger, 2012).

Ausgehend davon wurden die Aufgaben in mehreren Qualitätssicherungsschleifen überprüft. In Abbildung 7 sind die verschiedenen Schritte dieses Prozesses skizziert.

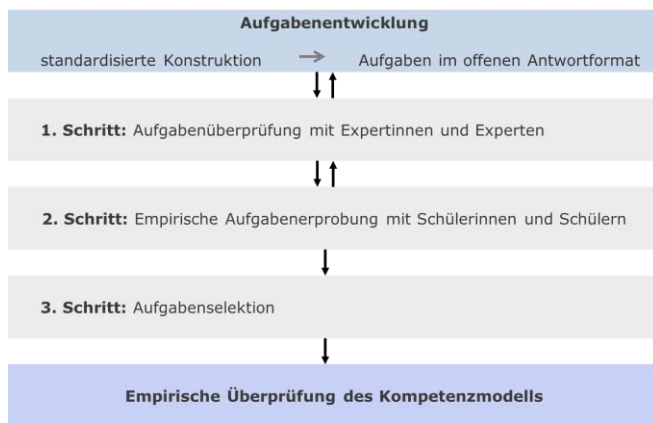


Abb. 7: Überblick über die Qualitätssicherungsschleifen (grau hinterlegt sind Schritte der Aufgabenüberprüfung und –selektion)

In einem ersten Schritt wurden die entwickelten Aufgaben mit sieben Expertinnen und Experten der Didaktik der Biologie diskutiert (Abb. 7; vgl. Neuhaus & Braun, 2007). Diese Personen zeichnen sich durch Erfahrungen mit der Zielgruppe und durch Kenntnisse des Kompetenzmodells der Modellkompetenz aus. Die Expertinnen und Experten beurteilten, inwiefern der Aufgabenstamm es ermöglichte, dass Lernende je nach ihrer Kompetenzausprägung plausibel auf allen drei Niveaus des Kompetenzmodells antworten würden (vgl. Kauertz et al., 2010). Hierfür wurden im Vorhinein Erwartungshorizonte pro Aufgabe formuliert und als Diskussionsgrundlage herangezogen. Darüber hinaus war zu bewerten, inwiefern die Informationen im Aufgabenstamm sprachlich und fachlich der Zielgruppe angemessen waren (vgl. Neuhaus & Braun, 2007). Auf der

Grundlage dieser Diskussionsergebnisse wurden die Aufgaben überarbeitet bzw. gegebenenfalls verworfen.

In einem zweiten Schritt erfolgte die empirische Erprobung der überarbeiteten Aufgaben mit Schülerinnen und Schülern (Abb. 7). Bei einer solchen Erprobung ist darauf zu achten, dass man eine Stichprobe aus der anvisierten Zielgruppe wählt, um Aussagen über die oben genannten Kriterien treffen zu können (Rost, 2004). Die Zielgruppe ist für diesen Test breit angelegt. Daher ist es sinnvoll, insbesondere die Extremgruppen (z. B. Realschule, 7. Jgst. und Gymnasium, 10. Jgst.) in die Teilstudien einzubinden. Infolgedessen nahmen an der Teilstudie 1 (insgesamt $N = 510$) Realschülerinnen und -schüler der siebten Jahrgangsstufe und Gymnasialschülerinnen und -schüler der zehnten Jahrgangsstufe teil. In Teilstudie 2 (insgesamt $N = 721$) umfasste die Stichprobe Realschülerinnen und -schüler der siebten und der neunten Jahrgangsstufe sowie Gymnasialschülerinnen und -schüler der achten und der neunten Jahrgangsstufe. Um den Bearbeitungsaufwand für die zu testenden Personen möglichst gering zu halten und ausführlich begründete Aussagen zu erhalten, wurden die Aufgaben auf verschiedene Testhefte verteilt. Jedes Testheft bestand aus mindestens drei und höchstens fünf Aufgaben. Für die Teilstudie 1 ($N = 25$ Aufgaben) lagen somit pro Aufgabe zwischen $N = 101$ und $N = 104$ Bearbeitungen und für die Teilstudie 2 ($N = 15$ Aufgaben) zwischen $N = 52$ und $N = 132$ Bearbeitungen pro Aufgabe vor. Die unterschiedliche Anzahl ist vor allem auf praktische Ursachen zurückzuführen. So sagten einige Klassen kurzfristig ihre Teilnahme an der Studie ab bzw. einige Lernende erhielten von ihren Eltern keine Genehmigung zur Teilnahme. Von weiteren Befragungen wurde allerdings abgesehen, weil bereits eine Sättigung der Daten erreicht war (vgl. Krüger & Riemeier, 2014), das heißt, dass bei der Auswertung der Ergebnisse der letzten Klassen keine neuen Hinweise auf Bearbeitungsschwierigkeiten identifiziert werden konnten.

Zur Prüfung der Plausibilität und Verständlichkeit des Aufgabenstamms wurden die Lernenden dazu aufgefordert, diejenigen Stellen in den Aufgaben zu kennzeichnen bzw. zu kommentieren, die sie nicht verstanden. Darüber hinaus protokollierten die Testleiterinnen und Testleiter während der Testungen alle aufkommenden Fragen der Lernenden sowie die darauf gegebenen Antworten. Über die Fragen und Kommentare der Lernenden hinaus bietet die Passung von Frageintention und Antwort einen Hinweis darauf, ob die Aufgaben ver-

ständig formuliert sind. So wurde bei der Auswertung der Aussagen der Lernenden darauf geachtet, ob diese die Informationen aus dem Aufgabenstamm nur wiederholten bzw. welche weiteren Hinweise auf das Verständnis aus den Aussagen der Lernenden zu identifizieren waren. Ob sich diese Aussagen auf die jeweilige Teilkompetenz bezogen (**eindeutige Modelleinordnung**, Kauertz et al., 2010), konnte ebenfalls über die Passung von Frageintention und Antwort geprüft werden.

Mit Hinblick auf das Kriterium der Repräsentation des Spektrums der jeweiligen Teilkompetenz (**Modellsättigung**, Kauertz et al., 2010) wurde überprüft, inwiefern über alle Aussagen der Lernenden hinweg die drei Niveaus der entsprechenden Teilkompetenz bedient wurden. Wenn Verständnisprobleme auftraten, wurden die jeweiligen Begriffe oder Abbildungen ersetzt oder Zusammenhänge anders dargestellt, und diese Aufgaben wurden dann erneut getestet. Zur Verbesserung der Aufgaben wurden sprachliche Vorschläge der Lernenden bzw. der Testleiterinnen und Testleiter genutzt, um eine bessere Anbindung an die Sprache der Lernenden zu gewährleisten (vgl. Neuhaus & Braun, 2007).

Nach Abschluss der empirischen Erprobung erfolgte in einem letzten Schritt die **Selektion** der Aufgaben für die empirische Überprüfung des Kompetenzmodells der Modellkompetenz. Hierfür wurde die prozentuale Verteilung der Aussagen der Lernenden auf die Niveaus berechnet. Darüber hinaus wurde betrachtet, wie viel Prozent der gesamten Aussagen pro Aufgabe auswertbar waren. Die Aufgaben, die auch nach einer Überarbeitung verstärkt Antworten auf einzelnen Niveaus induzierten oder nur von wenigen Lernenden beantwortet wurden, wurden nicht weiter verwendet (vgl. Grünkorn & Krüger, 2012; Grünkorn et al., 2011).

4.2 Ergebnisse der Operationalisierung

Die Ergebnisse hinsichtlich der beschriebenen Testkonstruktion sind in den folgenden zwei Publikationen veröffentlicht:

- Grünkorn, J., Upmeier zu Belzen, A. & Krüger, D. (2011). Design and test of open-ended tasks to evaluate a theoretical structure of model competence. In A. Yarden & G.S. Carvalho (Eds.), *Authenticity in Biology Education. Benefits and Challenges* (S. 53-65). Braga, Portugal: CIEC Universidade do Minho.
- Grünkorn, J. & Krüger, D. (2012). Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells der Modellkompetenz. In U. Harms & F.X. Bogner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (Band 5, S. 9-27). Innsbruck: Studienverlag.

4.2.1 Beitrag 1: Design and test of open-ended tasks to evaluate a theoretical structure of model competence

Grünkorn, J., Upmeier zu Belzen, A. & Krüger, D. (2011). Design and test of open-ended tasks to evaluate a theoretical structure of model competence. In A. Yarden & G.S. Carvalho (Eds.), *Authenticity in Biology Education. Benefits and Challenges* (S. 53-65). Braga, Portugal: CIEC Universidade do Minho.

Zusammenfassung

Der Artikel fokussiert auf die Ergebnisse hinsichtlich der Operationalisierung der Teilkompetenzen *Eigenschaften von Modellen* und *Alternative Modelle*. In zwei Teilstudien (TS1 und TS2) wurde der zentralen Forschungsfrage nachgegangen, inwiefern die Aussagen der Lernenden zu den entwickelten Aufgaben als Indikatoren der Teilkompetenzen *Eigenschaften von Modellen* und *Alternative Modelle* interpretiert werden können. Um dies zu überprüfen, sind folgende Detailfragen relevant:

- F1 Inwiefern lassen sich in den Aussagen und Kommentaren der Lernenden Verständnisschwierigkeiten identifizieren, die darauf hinweisen, dass die gewählten Itemstämme für Lernende nicht plausibel und/oder nicht verständlich formuliert sind?
- F2 Inwiefern sind die drei Niveaus der jeweiligen Teilkompetenz abgebildet?

Es wurden Aufgaben, welche die genannten Kriterien erfüllen, herangezogen, um Hinweise auf eine inhaltliche Überprüfung der beiden Teilkompetenzen zu geben. Dies war deshalb möglich, weil bereits im Rahmen von TS1 und TS2 verschiedene Perspektiven der Lernenden auf Modelle und auf den Modellbildungsprozess identifiziert worden waren:

F3 Inwieweit werden mit der theoretischen Struktur des Kompetenzmodells die Perspektiven von Schülerinnen und Schülern auf Modelle und auf die Modellbildung umfassend beschrieben?

Für die Teilkompetenz *Eigenschaften von Modellen* wurden fünf und für die Teilkompetenz *Alternative Modelle* zehn Aufgaben entwickelt, geprüft, überarbeitet und erneut getestet. Die Anzahl der Schülerinnen und Schüler aus Realschulen und Gymnasien, die Aufgaben zu diesen beiden Teilkompetenzen in TS1 (Realschule, 7. Jgst. und Gymnasium, 10. Jgst.) und TS2 (Realschule, 8. bis 10. Jgst. und Gymnasium, 9. bis 10. Jgst.) bearbeitet haben, beträgt $N = 912$ Schülerinnen und Schüler (insgesamt $N = 1\ 615$ Antworten von Lernenden).

Von insgesamt fünfzehn Aufgaben erfüllen sechs (jeweils drei Aufgaben pro Teilkompetenz) die im Vorhinein formulierten Kriterien, um die entsprechenden Antworten der Lernenden valide als Indikatoren der entsprechenden Teilkompetenz interpretieren zu können: Verständlichkeit und Plausibilität des Aufgabenstamms, Bezug zur jeweiligen Teilkompetenz und Repräsentation des Spektrums der entsprechenden Teilkompetenz.

Verständnis- und Interpretationsschwierigkeiten (F1) zeigten sich bei einzelnen Begriffen im Stimulus, die nach der Überarbeitung nicht mehr auftraten. So wiesen die Lernenden in ihren Aussagen darauf hin, dass sie mit dem Begriff *vergleichbar* im Stimulus der Teilkompetenz *Eigenschaften von Modellen* Verständnisschwierigkeiten hatten. Daher wurde der Stimulus in „Beschreibe, inwieweit das Modell des [Originals] so aussieht wie das [Original]. Begründe deine Antwort!“ geändert. Schwierigkeiten zeigten sich auch beim Stimulus für die Teilkompetenz *Alternative Modelle*, der zu offen und unpräzise formuliert war: „Begründe, warum es verschiedene [Original]-Modelle gibt!“ Dies führte einerseits dazu, dass 25 Prozent der Lernenden (126 von 510) alternative Modelle eines Originals auf die Annahme zurückführten, es gebe verschiedene Originale. Andererseits waren die Begründungen der Lernenden zum Teil zu

kurz und damit schwierig auszuwerten. Daher wurde in den Fachinformationen des Aufgabenstamms und im Stimulus eindeutig formuliert, dass es sich um Modelle nur eines Originals handelte. Weiterhin wurde der Stimulus um die **Aufforderung „Begründe deine Antwort!“ erweitert. Im Hinblick auf die gewählten Kontexte** erwiesen sich abstrakte Kontexte, beispielsweise die DNA-Struktur oder die Replikation, bei beiden Teilkompetenzen als ungeeignet, weil sie für die Lernenden der Zielgruppe zu schwer waren und nur wenige von ihnen darauf geantwortet hatten.

Die Analyse der Aussagen der Lernenden zeigte, dass sich diese auf die jeweilige Teilkompetenz beziehen und das Spektrum der Teilkompetenz abbilden (F2). Für die Teilkompetenz **Eigenschaften von Modellen** antwortete bei TS1 ($n = 509$; 5 Aufgaben) und TS2 ($n = 392$; 3 Aufgaben) ein Großteil der Lernenden auf Niveau I (TS1: 50 % und TS2: 68 %),⁷ gefolgt von Niveau II (16 % und 15 %) und Niveau III (4 % und 3 %).

Für die Teilkompetenz **Alternative Modelle** generierten in der TS1 ($n = 510$; 5 Aufgaben) 28 % der Lernenden Antworten auf Niveau I und 19 % auf Niveau II. In TS2 ($n = 204$; 5 Aufgaben) antworteten hingegen 34 % der Lernenden auf Niveau II und 27 % auf Niveau I. Niveau III erreichten bei beiden Teilstudien nur wenige Schülerinnen und Schüler (5 % bzw. 15 %).

Neben der Operationalisierung der Teilkompetenzen selbst wurde auch geprüft, inwiefern die Teilkompetenzen die Perspektiven der Schülerinnen und Schülern umfassend beschreiben (F3). Für die Teilkompetenz **Eigenschaften von Modellen** konnten keine zusätzlichen Perspektiven der Schülerinnen und Schülern identifiziert werden.

In den Antworten der Lernenden zur Teilkompetenz **Alternative Modelle** konnte hingegen eine Kategorie identifiziert werden, die sich keinem der drei Niveaus zuordnen ließ. Es handelt sich um die Kategorie **Verschiedene Modelle zu unterschiedlichen Originalen**. Diese Kategorie ist dadurch gekennzeichnet, dass die Lernenden verschiedene Modelle mit unterschiedlichen Originalen begründen, obwohl deutlich hervorgehoben wurde, dass es Modelle **eines** Originals sind. Diese Lernenden schienen also zu bezweifeln, dass es alternative Modelle eines Originals gab.

⁷ Die Werte repräsentieren das höchst Niveau der Lernenden. Die Cohens-Kappa-Werte (κ) für die Teilstudie 2 sind im Artikel nicht aufgeführt. Diese betragen bei den Teilkompetenzen **Eigenschaften von Modellen** und **Alternative Modelle** jeweils $\kappa = .8$.

Darüber hinaus weisen die Ergebnisse darauf hin, dass die Lernenden ein alternatives Verständnis des Begriffs *Modell* hatten. Das Spektrum reicht von der Ablehnung gewisser Repräsentationen (z. B. Zeichnungen) als Modelle bis hin zur Akzeptanz aller Repräsentationen als Modelle.

4.2.2 Beitrag 2: Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells der Modellkompetenz

Grünkorn, J. & Krüger, D. (2012). Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells der Modellkompetenz. In U. Harms & F.X. Bogner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (Band 5, S. 9-27). Innsbruck: Studienverlag.

Zusammenfassung

Der Artikel beschreibt die Ergebnisse der Operationalisierung der Teilkompetenzen *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen*. Das Ziel war, Aufgaben für diese Teilkompetenzen zu entwickeln, deren Antworten als Indikatoren von Modellkompetenz interpretiert werden können. Folgende Fragestellungen wurden in zwei Teilstudien (TS1 und TS2) untersucht:

- F1 Inwiefern sind die gewählten Itemstämme für Lernende plausibel und verständlich formuliert?
- F2 Inwiefern lassen sich die Antworten der Lernenden auf die jeweilige Teilkompetenz beziehen?
- F3 Inwiefern bilden sich die drei Niveaus der jeweiligen Teilkompetenz in den Antworten der Lernenden ab?

Auch bei diesem Beitrag wurden Aufgaben, welche die genannten Kriterien erfüllen, herangezogen, um Hinweise auf eine inhaltliche Überprüfung der Teilkompetenzen *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen* zu geben:

- F4 Inwieweit werden mit der theoretischen Struktur des Kompetenzmodells die Perspektiven von Schülerinnen und Schülern auf Modelle und auf die Modellbildung umfassend beschrieben?

Für die Teilkompetenzen *Zweck von Modellen* und *Testen von Modellen* wurden acht und für die Teilkompetenz *Ändern von Modellen* neun Aufgaben entwickelt, geprüft, überarbeitet und erneut getestet. In TS1 (Realschule, 7. Jgst. und Gymnasium, 10. Jgst.) und TS2 (Realschule, 7. und 9. Jgst. sowie Gymnasium, 8. bis 9. Jgst.) nahmen insgesamt $N = 1\,011$ Lernende (insgesamt $N = 2\,277$ Bearbeitungen) an der empirischen Erprobung der Aufgaben teil.

Die Ergebnisse der Untersuchungen zeigten, dass von insgesamt 25 Aufgaben neun Aufgaben (jeweils drei Aufgaben pro Teilkompetenz) die genannten Kriterien erfüllten und es daher zulässig ist, von diesen Aufgaben auf die Modellkompetenz der Lernenden zu schließen. Im Hinblick auf die Verständlichkeit und Plausibilität (F1) sowie den Bezug zur jeweiligen Teilkompetenz (F2) traten bei der Teilkompetenz *Zweck von Modellen* vor allem Schwierigkeiten hinsichtlich der Fachinformationen im Aufgabenstamm auf. Diese wurden von den Lernenden zum Teil vollständig wiederholt oder waren zu stark lenkend. Daher mussten für diese Teilkompetenz sämtliche Fachinformationen im Aufgabenstamm herausgenommen werden. Diese Veränderung konnte allerdings nur bei denjenigen Kontexten (z. B. Wald) vorgenommen werden, die auch ohne Fachinformationen verständlich sind. Andere Kontexte, beispielsweise Enzyme oder Schwimmblase, konnten hingegen nicht weiterbenutzt werden. Diese Aufgaben wären sonst im Hinblick auf das Fachwissen, das für das Verständnis des Modells benötigt wird, nicht verstanden worden.

Bei den Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* zeigten sich Schwierigkeiten mit den Stimuli. So löste der Stimulus der Teilkompetenz *Testen von Modellen* „Begründe, wie man überprüfen kann, ob das dargestellte Modell tauglich ist bzw. seinen Zweck erfüllt!“ zunächst Antworten aus, die sich nicht deutlich genug auf die Teilkompetenz bezogen. Nachdem im Stimulus die Begriffe „wie“ und „überprüfen“ hervorgehoben wurden, trat diese Schwierigkeit nicht mehr auf.

Hinsichtlich der dritten Frage (F3) repräsentierten die Aussagen der Lernenden das gesamte Spektrum der jeweiligen Teilkompetenz. Für die Teilkompetenz *Zweck von Modellen* verteilen sich die Lernenden in TS1 ($n = 513$; 5 Aufgaben) und TS2 ($n = 215$; 4 Aufgaben; $\kappa = .84$) wie folgt: Niveau II

(TS1: 56 % und TS2: 43 %),⁸ gefolgt von Niveau I (12 % und 25 %) und Niveau III (2 % und 23 %).

Die prozentuale Verteilung der Lernenden für die Teilkompetenz **Testen von Modellen** in TS1 ($n = 509$; 5 Aufgaben) und 2 ($n = 245$; 4 Aufgaben; $\kappa = .81$) zeigte, dass ein Großteil der Lernenden auf Niveau II (34 % und 53 %) antwortete. Niveau I erreichten 10 % bzw. 20 % und Niveau III (5 % und 9 %) war am seltensten vertreten.

Bei der Teilkompetenz **Ändern von Modellen** antworteten die meisten Lernenden in TS1 ($n = 510$; 5 Aufgaben) und TS2 ($n = 285$; 5 Aufgaben; $\kappa = .81$) auf Niveau II (53 % und 45 %). Deutlich weniger Schüler antworteten auf Niveau I (10 % und 13 %). Niveau III (0,2 % und 2 %) wurde selten erreicht.

Zusätzlich konnten neben der Überprüfung der Aufgaben auch Hinweise auf die inhaltliche Struktur des Kompetenzmodells gefunden werden (F4): Für die Teilkompetenz **Zweck von Modellen** konnten keine weiteren Perspektive der Lernenden beschrieben werden. Für die Teilkompetenzen **Testen von Modellen** und **Ändern von Modellen** wurde jeweils eine Perspektive identifiziert, die im Kompetenzmodell bis dahin nicht abgebildet war. Hier äußerten die Lernenden ihre Ratlosigkeit darüber, warum Modelle getestet (1 %) oder geändert werden (4 %) müssen. Dies formulierten die Lernenden, obwohl – auch aus Sicht der übrigen Befragten – mit den gezeigten Modellen eine Anwendung bzw. Testung möglich und plausibel ist.

⁸ Die Werte repräsentieren das höchste Niveau der Lernenden.

4.3 Zusammenfassende Diskussion

Mit der theoriegeleiteten Testentwicklung (vgl. S. 54ff. dieser Arbeit) und den empirischen Befunden aus TS1 und TS2 (Grünkorn & Krüger, 2012; Grünkorn et al., 2011) wurden Argumente dafür herausgearbeitet, inwiefern die folgenden Kriterien erfüllt sind:

- **Plausibilität und Verständlichkeit** des Aufgabenstamms,
- **Bezug** der Aussagen der Lernenden zur jeweiligen Teilkompetenz und
- **Repräsentation** der drei Niveaus der jeweiligen Teilkompetenz in den Aussagen der Lernenden.

Daraus ergibt sich abschließend ein Gesamtbild hinsichtlich der Aussagekraft des Tests zur Modellkompetenz (vgl. Kane, 2001, 2013).

Plausibilität und Verständlichkeit

Ausgehend von den Qualitätssicherungsschleifen im Rahmen der Testkonstruktion und den Ergebnissen aus der empirischen Erprobung mit Lernenden der Zielgruppe ist anzunehmen, dass die Aufgabenstämme plausibel und verständlich sind. Dies ist ein zulässiger Schluss, weil vor der empirischen Erprobung mit Lernenden die nachstehenden Faktoren berücksichtigt und mit Expertinnen und Experten diskutiert (vgl. Neuhaus & Braun, 2007) wurden:

- **Standardisierte Struktur** der Aufgaben (vgl. Jonkisz et al. 2012; Rost, 2004; Terzer et al., 2013)
- **Kontextauswahl** (vgl. Hammann, 2006)
- Berücksichtigung weiterer schwierigkeiterzeugender Merkmale, wie **Aufgabenformat** (vgl. Hartig & Klieme, 2006; Martinez, 1991), **formale Merkmale** (vgl. Jonkisz et al., 2012; Kauertz et al., 2010; Prenzel et al., 2002; Rost, 2004) und **Fachwissen** (vgl. Kauertz et al., 2010; Prenzel et al., 2002).

Wichtige Hinweise sind zudem in den Konstruktionsanleitungen pro Teilkompetenz festgehalten (Anhang 3). Andererseits zeigen die Ergebnisse der empirischen Erprobung mit den Lernenden zusätzliche Verständnis- aber auch Interpretationsschwierigkeiten, die nach der Optimierung nicht mehr auftraten. Dies äußerte sich vor allem darin, dass in Teilstudie 2 wenige bis keine Nachfragen während der Testung gestellt wurden sowie die Passung von Frageintention

und Antwort verbessert wurde und damit vermehrt Aussagen der jeweiligen Teilkompetenz zugeordnet werden konnten. Zudem kommentierten nur sehr wenige Lernende, dass sie Abschnitte oder Abbildungen nicht verstanden hatten (Grünkorn & Krüger, 2012; Grünkorn et al., 2011). Im Hinblick auf die Verbesserung der Verständlichkeit und Plausibilität waren bei Fragen der Lernenden die Ersatzformulierungen der Testleiterinnen und Testleiter besonders hilfreich. Hier zeigte die Reaktion der Lernenden sogleich, ob sie die Umformulierung besser verstanden.

Die Verbesserung der Verständlichkeit und Plausibilität kann auch in den unterschiedlichen Stichproben von TS1 und TS2 begründet liegen, so dass Unterschiede in der Bearbeitung auf interindividuelle Unterschiede zurückzuführen sein könnten. Diese Erklärung ist allerdings nicht stichhaltig, denn die Variable *Schulart* (gleiche Verteilung auf Realschule und Gymnasium) wurde konstant gehalten, es wurden keine besonderen Klassen (z. B. Schnellläuferklassen) eingebunden und es wurde eine vergleichsweise große Stichprobengröße in beiden Teilstudien (TS1 $N = 510$; TS2, $N = 721$) gewählt. Vielmehr kann angenommen werden, dass die Überarbeitung der Aufgaben zu einem besseren Verständnis der Aufgaben geführt hat.

Bezug zur Teilkompetenz und Repräsentation des Spektrums der Teilkompetenz

Aufgrund der präsentierten Analysen ist anzunehmen, dass die fünfzehn Aufgaben Antworten zulassen, die sich auf die entsprechende Teilkompetenz beziehen, und die Antworten das gesamte Spektrum der jeweiligen Teilkompetenz abbilden. Diese Schlussfolgerung ist daher zulässig, weil die Aufgaben theoriebasiert mittels Konstruktionsanleitungen (Anhang 3) entwickelt wurden, wobei in jedem Schritt klare Bezüge zu den drei Niveaus der jeweiligen Teilkompetenz hergestellt wurden (vgl. Hartig et al., 2012; Kane 2001). Weiterhin zeigen die Ergebnisse der empirischen Erprobung, dass die Aufgaben Antworten im gesamten Spektrum der Teilkompetenz auslösen (Grünkorn & Krüger, 2012; Grünkorn et al., 2011). Dass die jeweilige Teilkompetenz adäquat repräsentiert war, wurde dadurch gewährleistet, dass die entwickelten Kategorien in einem ersten Durchgang theoriegeleitet den drei Niveaus zugeordnet und diese Zuordnung mit Expertinnen und Experten diskutiert wurde. Um darüber hinaus die Zuordnung der Aussagen der Lernenden zu den Kategorien zu

objektivieren, wurde die Übereinstimmung zwischen jeweils zwei Kodierenden bestimmt. Diese lag zwischen $\kappa = .80$ bzw. $\kappa = .84$ und ist als gut einzuschätzen (Bortz & Döring, 2006). Somit lassen sich die Aussagen auf eine Kategorie bzw. ein Niveau innerhalb einer Teilkompetenz beziehen und repräsentieren das gesamte Spektrum der jeweiligen Teilkompetenz (vgl. Kauertz et al., 2010).

Die prozentuale Verteilung der Antworten der Lernenden auf die drei Niveaus der jeweiligen Teilkompetenz deutet zunächst auf einen schweren Test hin. Dies könnte unter anderem durch das Aufgabenformat zu erklären sein (Hartig & Klieme, 2006; Martinez, 1991). Allerdings ist zu berücksichtigen, dass die identifizierten Kategorien sowie die Verteilung mit denen anderer Studien vergleichbar sind (u. a. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Trier & Upmeier zu Belzen, 2009). Dies spricht dafür, dass sich die Aufgabenschwierigkeiten vor allem durch die jeweilige Teilkompetenz erklären lassen.

Die Ergebnisse der Teilstudien zeigen zudem, dass über die drei Niveaus der jeweiligen Teilkompetenz hinaus zusätzliche Perspektiven von Schülerinnen und Schülern auf Modellen und auf den Modellbildungsprozess erfasst wurden. Zusätzliche Kategorien konnten für die Teilkompetenz *Alternative Modelle, Testen von Modellen* und *Ändern von Modellen* identifiziert werden (Grünkorn & Krüger, 2012; Grünkorn et al., 2011). Die Ergebnisse sprechen dafür, dass sich die Aufgaben auch für eine inhaltliche Überprüfung des Kompetenzmodells eignen.

Einschränkungen der Aussagekraft des Tests

Auch wenn der Test für die Erfassung von Modellkompetenz im Rahmen dieser Arbeit geeignet ist, bleibt seine Aussagekraft – wie bei jeder empirischen Untersuchung – eingeschränkt. So fokussiert der entwickelte Test auf die *kognitive Komponente* von Modellkompetenz, weswegen über die volitionale, die motivationale und die soziale Komponente von Modellkompetenz keine Aussage getroffen werden kann. Zudem werden keine Handlungen in den Blick genommen. Für diese Komponenten bedarf es zusätzlicher Tests, die für die Messung dieser Komponenten geeignet sind (z. B. Hands-on-Aufgaben, Orsenne & Upmeier zu Belzen, 2012). Auf der Basis unterschiedlicher Tests können Zusammenhänge zwischen den verschiedenen Komponenten untersucht und Hinwei-

se darauf abgeleitet werden, wie sich die Modellkompetenz fördern lässt (vgl. Weinert, 1999, 2001).

Außerdem ist der Test für eine bestimmte **Zielgruppe** (Realschule und Gymnasium, 7. bis 10. Jgst.) konstruiert, in dieser getestet und nach den genannten Kriterien überprüft worden. Eine Verwendung dieses Tests bei anderen Zielgruppen kann ohne weitere Prüfung nicht erfolgen (Jonkisz et al., 2012; Rost 2004). So könnte der Test beispielweise Deckeneffekte (**ceiling effects**) oder Bodeneffekte (**floor effects**) erzeugen, wenn er für andere Zielgruppen zu leicht oder zu schwer ist (Rost, 2004).

Auch die Übertragbarkeit auf andere Domänen oder Fächer, wie Physik und Chemie, kann nicht ohne weitere Prüfung angenommen werden. Das liegt darin begründet, dass Kompetenzen in einer bestimmten Domäne erworben und auch in dieser angewandt werden (vgl. Hartig & Klieme, 2006; Klieme & Steinert, 2004). Der Test ist für die **Domäne Biologie** konstruiert und beinhaltet Aufgaben, die verschiedene biologische Situationen beschreiben. Zudem wird in den naturwissenschaftlichen Fächern der Begriff **Modell** unterschiedlich verwendet (Beerenwinkel & Parchmann, 2008; Mikelskis-Seifert & Fischler, 2003). In der Biologie sind vielfach gegenständliche Struktur- und Funktionsmodelle oder dynamische Modelle (z. B. Ökosystem) im Einsatz, wohingegen in Physik und Chemie vor allem abstrakte oder mathematische Modelle vorherrschen. Daher lassen sich die in der Domäne Biologie gemessenen Kompetenzen nicht verallgemeinern (Hartig & Klieme, 2006; Klieme & Steinert, 2004).

Eine weitere Einschränkung der Aussagekraft des Tests ist durch die hier gewählte Kompetenzdefinition von Klieme und Leutner (2006) und den damit einhergehenden **kontextabhängigen Ansatz** begründet. Dies bedeutet, dass die verschiedenen Aufgaben in unterschiedliche konkrete Situationen eingebettet sind. Die Verwendung unterschiedlicher Kontexte beeinflusst allerdings das Antwortverhalten der Lernenden und ist somit schwierigkeiterzeugend (vgl. u. a. Krell, Upmeyer zu Belzen et al., 2012b, 2014c; Schecker & Parchmann, 2006; Sins et al., 2009). Der Einsatz dieses Tests zielt darauf ab, verschiedene Perspektiven von Schülerinnen und Schülern auf Modelle und auf den Modellbildungsprozess zu identifizieren und, darauf aufbauend, das Kompetenzmodell inhaltlich zu überprüfen. Vor diesem Hintergrund ist eine systematische Variation des Kontexts nicht notwendig. Infolge des **Einsatzbereichs** bzw. des **Geltungsbereichs** des entwickelten Tests ist dieser zunächst nur zur Überprüfung

des Kompetenzmodells der Modellkompetenz, aber nicht in der Individualdiagnostik einsetzbar (vgl. Klieme, Avenarius et al., 2007; Leutner et al., 2007; Terzer et al., 2013). Für den letzteren Einsatzbereich sind mehr Aufgaben und eine gezielte Variation des Kontextes zu empfehlen. Gogolin und Krüger (*in Druck*), die Modellkompetenz unter anderem mit den hier geprüften Kontexten diagnostizieren, berücksichtigen diesen Faktor, indem sie die Einteilung der Modelle nach den Kriterien von Krell, Upmeyer zu Belzen und Krüger (2014b) in *gegenständlich* und *abstrakt* vornehmen.

4.4 Fazit

Das erste Ziel dieser Forschungsarbeit – die Operationalisierung des Kompetenzmodells der Modellkompetenz mit Aufgaben im offenen Antwortformat – wurde in fünfzehn Aufgaben (drei Aufgaben pro Teilkompetenz) realisiert, welche die kognitive Komponente von Modellkompetenz erfassen (Grünkorn & Krüger, 2012; Grünkorn et al., 2011). Die Aufgabenstämme sind verständlich und plausibel, der Bezug zur jeweiligen Teilkompetenz ist nachvollziehbar und das Spektrum der jeweiligen Teilkompetenz ist angemessen repräsentiert. Deshalb ist davon auszugehen, dass die Aufgaben Antworten zulassen, die als Indikatoren von Modellkompetenz interpretiert werden können (vgl. Hartig et al., 2012; Kane 2001, 2013) und somit zur Überprüfung der inhaltlichen Struktur des Kompetenzmodells für die Domäne Biologie geeignet sind.

5 Analyse und Beschreibung von Modellkompetenz

Die Befunde aus den beschriebenen empirischen Studien zu den Vorstellungen der Schülerinnen und Schüler von Modellen und dem Modellbildungsprozess zeigen, dass die Lernenden Modelle vorwiegend aus einer medialen Sicht betrachten (z. B. Grosslight et al., 1991; Grünkorn & Krüger, 2012; Grünkorn et al., 2011; Krell, Upmeier zu Belzen et al. 2012b; Schwarz et al., 2009; Schwarz & White, 2005; Treagust et al., 2002; Terzer, 2013; Trier & Upmeier zu Belzen, 2009). Für die Entwicklung einer umfassenden Modellkompetenz ist sowohl die Sicht auf Modelle als Medien zur Veranschaulichung als auch die Sicht auf Modelle als Methode zur Erkenntnisgewinnung wichtig. Somit ist es notwendig, die methodische Perspektive der Schülerinnen und Schüler zu fördern. Um die Lehrkräfte bei der Entwicklung von Fördermaßnahmen zu unterstützen, sind empirische Befunde über die Perspektivenvielfalt der Lernenden, über die Strukturierung von Modellkompetenz und über Besonderheiten in den Antwortstrukturen der Lernenden hilfreich. Aus diesen Gründen ist die Analyse und strukturierende Beschreibung der Modellkompetenz von Schülerinnen und Schülern der Sekundarstufe I Schwerpunkt dieses Kapitels (Abb. 8).

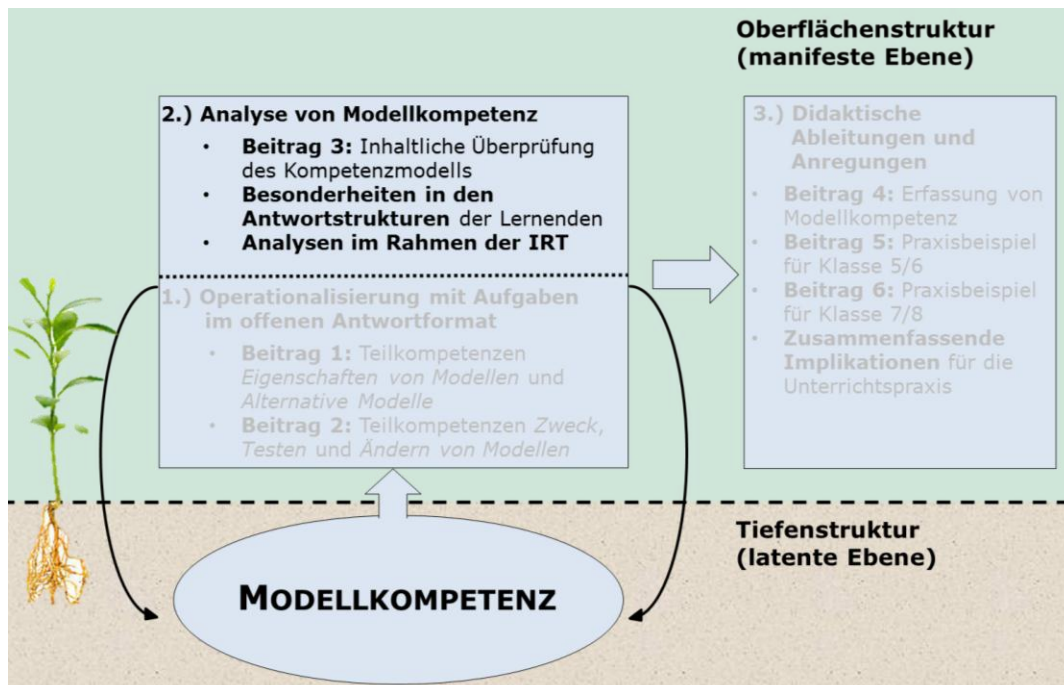


Abb. 8: Zweites Ziel der Forschungsarbeit: Analyse und Beschreibung von Modellkompetenz (Grafik nach Meyer, 2012)

Für die Analyse und Beschreibung von Modellkompetenz wurde eine Querschnittsstudie mit 1 177 Berliner Gymnasiastinnen und Gymnasiasten der siebten bis zehnten Jahrgangsstufe gewählt (vgl. S. 86ff. dieser Arbeit). Die so gewonnenen Daten wurden nach der qualitativen Inhaltsanalyse (Mayring, 2010) ausgewertet, die Aufschluss über die Vielfalt an Perspektiven auf Modelle und auf den Modellbildungsprozess der Schülerinnen und Schüler gibt. Darauf aufbauend wurde geprüft, inwieweit das Kompetenzmodell der Modellkompetenz die Perspektiven von Lernenden hinreichend beschreibt (Grünkorn, Upmeier zu Belzen et al., 2014). Diese inhaltliche Analyse stellt – in Verbindung mit den Ergebnissen von Krell (2012; Krell & Krüger, 2011) und Terzer (2013) – einen weiteren Baustein der empirischen Überprüfung des Kompetenzmodells dar (vgl. S. 100ff. dieser Arbeit). Ergänzend dazu wurde anhand dieser Daten untersucht, ob sich bei den Lernenden dieser Stichprobe Besonderheiten in den Antwortstrukturen identifizieren lassen, um Erklärungsansätze für die Perspektiven von Schülerinnen und Schülern zu liefern (vgl. S. 104ff. dieser Arbeit).

Über diese Analysen hinaus, die eine evidenzbasierte Entwicklung von Fördermaßnahmen ermöglichen und damit ein weiteres Forschungsfeld eröffnen, wurde geprüft, inwieweit der entwickelte Leistungstest auch in einem probabilistischen Ansatz einsetzbar ist (vgl. S. 111ff. dieser Arbeit). Dies ist insbesondere im Hinblick auf diejenigen Projekte zur Entwicklung (Patzke et al., *in Druck*) und zur Individualdiagnose von Modellkompetenz bei Schülerinnen und Schülern (Gogolin & Krüger, *in Druck*) sinnvoll, die mit dem hier entwickelten Leistungstest einen solchen Ansatz verfolgen.

5.1 Untersuchungsdesign und Methodik

Abbildung 9 liefert einen Überblick über die methodische Vorgehensweise der Querschnittsstudie zur Analyse und Beschreibung von Modellkompetenz: Das Datenmaterial wurde hierfür zunächst beschrieben, dann aufbereitet und schließlich inhaltsanalytisch ausgewertet, um im Anschluss daran statistische Analysen durchzuführen.

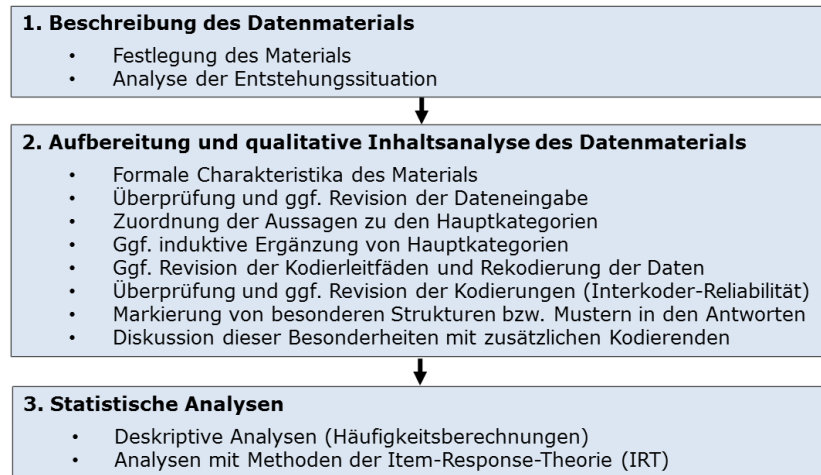


Abb. 9: Methodische Vorgehensweise bei der Analyse von Modellkompetenz (vgl. Gläser-Zikuda, 2005; Gropengießer, 2005; Mayring, 2010)

5.1.1 Beschreibung des Datenmaterials

Die Stichprobe orientierte sich an der festgelegten Zielgruppe (Lernende der 7. bis 10. Jgst.), die mit Blick auf die Bildungsstandards (KMK, 2005) ausgewählt wurde. Im Gegensatz zu den Teilstudien zur Operationalisierung, in die Lernende aus Realschulen und Gymnasien einbezogen waren (vgl. S. 70ff. dieser Arbeit), entstammt die Stichprobe dieser Untersuchung nur dem Gymnasium. Der Grund hierfür war die Reform des Schulsystems in Berlin im Schuljahr 2010/2011, in der die Haupt-, die Real- und die Gesamtschulen zu integrierten Sekundarschulen vereinigt wurden. Da der Test nicht für Hauptschülerinnen und Hauptschüler geprüft wurde, konnte er nicht in integrierten Sekundarschulen eingesetzt werden, denn man hätte hier nicht verlässlich von den Testergebnissen auf die Kompetenzen der jeweiligen Personen schließen können (vgl. Jonkisz et al., 2012; Kane, 2001).

Die Querschnittsstudie umfasste 1 177 Berliner Gymnasiastinnen und Gymnasiasten der siebten bis zehnten Jahrgangsstufe aus 48 verschiedenen Schulklassen (Tab. 4). Unterschiedliche Jahrgangsstufen wurden deshalb gewählt, um das Antwortspektrum in der Breite zu erfassen. Mit den unterschiedlichen Schulklassen und den verschiedenen Jahrgangsstufen liegt eine geschichtete bzw. stratifizierte Klumpenstichprobe vor (Bortz & Döring, 2006; Zöfel, 2003). Diese Querschnittsstudie erlaubt Aussagen über die Struktur von Modellkompetenz auf Populationsebene (Klassen- oder Schulebene), jedoch nicht über die individuelle Kompetenzentwicklung.

Tab. 4: Angaben zur Stichprobe der Studie zur Analyse von Modellkompetenz (vgl. Grünkorn, Upmeyer zu Belzen & Krüger, 2014)

Jgst.	n	Geschlechterverteilung			Alter				
		♀	♂	k. A.	Min.	Max.	M	SD	k. A.
7	313	160	150	3	11	14	12.70	0.57	5
8	293	143	149	1	12	16	13.83	0.62	4
9	283	166	113	4	13	17	14.67	0.62	2
10	288	157	130	1	14	19	15.95	0.69	1

Anmerkung. k. A. = keine Angaben

Die Teilnahme an der Studie war freiwillig und anonym. Die teilnehmenden Schulen wurden über das Ziel der Studie informiert, und es wurden Einverständniserklärungen sowohl von den Schulleitungen als auch von den Eltern eingeholt. Die Testungen fanden innerhalb der regulären Unterrichtszeit statt und wurden von der Autorin selbst oder von Testleiterinnen und Testleitern durchgeführt. Zur Objektivierung der Testdurchführung wurden die Testungen protokolliert und nach einem Testmanual durchgeführt (Anhang 4; vgl. Jonkisz et al., 2012; Neuhaus & Braun, 2007; Rost, 2004). Den Lernenden, die den Test schnell bearbeiteten, wurden außerdem Rätselaufgaben ausgeteilt, um Störungen während der Testung zu vermeiden.

Die Datenerhebung erfolgte gemeinsam mit Krell et al. (2014b) und Terzer (2013), so dass die Schülerinnen und Schüler in getrennten Testheften neben Aufgaben im offenen Antwortformat auch FC-Items (Krell et al., 2014b) und MC-Items (Terzer, 2013) bearbeiteten. Die Aufgaben im offenen Format wurden den Lernenden zuerst vorgelegt, um Effekte der vorgegebenen Antwortmöglichkeiten auf die freien Antworten zu vermeiden. Darüber hinaus wurden Kontrollvariablen wie Jahrgangsstufe, Alter und Geschlecht erfasst.

Um die Testzeit und damit die Beanspruchung für die Schülerinnen und Schüler möglichst gering zu gehalten, wurden die 15 Aufgaben im offenen Antwortformat auf verschiedene Testhefte verteilt. Dies bedeutet, dass die Schülerinnen und Schüler nicht alle Testaufgaben bearbeiteten, sondern nur jeweils drei Aufgaben aus dem Aufgabenpool vorgelegt bekamen. Die Anzahl von drei Aufgaben hatte sich bereits in den Teilstudien 1 und 2 bewährt (vgl. Grünkorn et al., 2011; Grünkorn & Krüger, 2012). Es wurde ein *balanced incomplete block design* (BIBD) genutzt (Colbourn & Dinitz, 1996; Frey, Hartig & Rupp, 2009; Giesbrecht & Gumpertz, 2004; Tab. 5). Dieser Multi-Matrix-Design Typ (MMX-Design, Carstensen, Frey, Walter & Knoll, 2007) ist dadurch charakterisiert, dass keine Aufgabe mehr als einmal pro Testheft und jede Aufgabe an jeder

Position im Testheft gleich oft vorkommt (vgl. Frey et al., 2009). Auf diese Weise können mögliche Veränderungen der Aufgabenschwierigkeit aufgrund der Platzierung im Test ausgemittelt werden (*Positionseffekt*, Rost, 2004). Die Aufgaben wurden auf 35 Testhefte verteilt, so dass pro Aufgabe etwa 235 schriftliche Antworten von Schülerinnen und Schülern vorlagen (insgesamt 3531 Antworten). Jede Aufgabe kam sieben Mal vor und wurde mit unterschiedlichen anderen Aufgaben kombiniert. Die Verteilung der Testhefte in den Schulklassen erfolgte in einem Spiraling-Verfahren, so dass die geschichtete Klumpenstichprobe gemischt wurde (Frey et al., 2009).

Tab. 5: Multi-Matrix-Design – (15, 3, 1)-BIBD (Colbourn & Dinitz, 1996) zur Analyse von Modellkompetenz (vgl. Terzer, 2013)

Position im Testheft	Testheftnummer																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Position 1	Z1	T1	EM2	EM1	Ä2	EM1	EM1	A1	Ä1	Z2	A1	A3	Ä3	Z1	Ä1	Z2	Z1
Position 2	A1	EM1	A2	Z2	EM1	Z3	T3	T1	A1	Ä2	T2	A1	Z3	A2	Z1	EM3	T2
Position 3	EM1	Ä1	EM1	T2	EM3	A3	Ä3	EM2	A2	A1	EM3	T3	A1	T1	EM2	Z1	Ä2

Anmerkung. EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen. Die Zahlen kennzeichnen die Nummer der Aufgabe, zum Beispiel Z1 = Aufgabe 1 in der Teilkompetenz *Zweck von Modellen*

Tab. 5: (Fortsetzung)

Position im Testheft	Testheftnummer																	
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Position 1	Ä3	Z3	T1	T2	T3	T1	Z3	A3	Ä2	EM3	EM2	T2	Ä2	EM3	A2	T3	A2	A2
Position 2	Z1	T3	A3	T1	Ä2	EM3	Ä1	Ä1	Ä3	Ä1	Z2	EM2	A3	EM2	Ä3	T2	Z3	EM3
Position 3	A3	Z1	Z2	Z3	T1	Ä3	Z2	T2	Ä1	T3	T3	Ä3	EM2	Z3	Z2	A2	Ä2	A3

Anmerkung. EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen. Die Zahlen kennzeichnen die Nummer der Aufgabe, zum Beispiel Z1 = Aufgabe 1 in der Teilkompetenz *Zweck von Modellen*

5.1.2 Aufbereitung und qualitative Inhaltsanalyse des Datenmaterials

Die schriftlichen Aussagen mussten für die Analysen aufbereitet werden. Die hierbei genutzte Vorgehensweise war analog zu der in den Teilstudien 1 und 2 zur Operationalisierung und ist dort bereits beschrieben (vgl. S. 58ff. dieser Arbeit). Es wurden alle Antworten aufbereitet; eine Auswahl wurde nicht getroffen.

Nach Abschluss der Aufbereitung erfolgte eine qualitative Inhaltsanalyse des Datenmaterials nach Mayring (2010) mit Unterstützung der Software MAXQDA

Version 10 (vgl. Teilstudie 2, S. 58ff. dieser Arbeit). Das bedeutet, dass die Antworten der Schülerinnen und Schüler mithilfe der in den Teilstudien 1 und 2 entwickelten Kodierleitfäden (Anhang 1)⁹ den theoriegeleiteten Hauptkategorien zugeordnet wurden. Für Aussagen, die nicht den bereits bestehenden Hauptkategorien zugeordnet werden konnten, wurden neue Kategorien definiert (vgl. Hammann & Jördens, 2014). Diese neuen Kategorien wurden zu Hauptkategorien zusammengefasst und unter Berücksichtigung der jeweiligen Definition der Teilkompetenz, der Beschreibung der Niveaus (Upmeier zu Belzen & Krüger, 2010) und der medialen und methodischen Sichtweisen auf Modelle (Gilbert, 1991; Mahr, 2009) ergänzt. Davon ausgehend wurden der Kodierleitfaden überarbeitet und die Daten rekodiert.

Alle Kodierungen wurden von der Autorin sowie drei weiteren Kodierenden durchgeführt. Die Kodierenden waren mit dem Kompetenzmodell der Modellkompetenz, den Kodierregeln und dem Kodierschema vertraut. Ungefähr die Hälfte des Datenmaterials wurde doppelt kodiert. Unterschiede in den Zuordnungen bzw. neue Kategorien wurden diskutiert, bis ein Konsens erreicht wurde (*konsensuelle Validierung*, Flick, 2002). Um die Übereinstimmung zwischen den Kodierenden zu bestimmen, wurden Berechnungen von Cohens Kappa (κ , Interkoder-Reliabilitäten) vorgenommen (Bortz & Döring, 2006). Die Interkoder-Reliabilitäten lagen bei dieser Studie zwischen $\kappa = .81$ und $\kappa = .90$ und sind als gut einzuschätzen (Bortz & Döring, 2006).

Die Ergebnisse dieser Analyse wurden zur Ausdifferenzierung der einzelnen Niveaus, zur inhaltlichen Überprüfung der theoretischen Struktur (Grünkorn, Upmeier zu Belzen et al., 2014) und zur Beschreibung von Besonderheiten in den Antwortstrukturen der Lernenden herangezogen (vgl. S. 104ff. dieser Arbeit). Für die Beschreibung von Besonderheiten wurden ergänzend auffällige Strukturen bzw. Muster in den Antworten markiert und mit den zusätzlichen drei Kodierenden diskutiert (vgl. Mayring, 2010). Solche Besonderheiten erlauben ein vertieftes Verständnis der Sichtweisen von Lernenden auf Modelle und stellen darüber hinaus Erklärungsansätze für die überwiegend mediale Perspektive auf Modelle bereit.

⁹ In den Kodierleitfäden wird keine Unterscheidung in Kategorie und Hauptkategorie vorgenommen. Wird von einem Kode gesprochen, ist damit stets die Hauptkategorie gemeint.

5.1.3 Statistische Analysen

Auf der Basis der überprüften Kodierungen wurde eine Datenmatrix für die statistischen Analysen erstellt, um, darauf aufbauend, Häufigkeiten zu berechnen und Analysen mit Hilfe der Item-Response-Theorie (IRT) durchzuführen (vgl. Gläser-Zikuda, 2005). Diese Datenmatrix enthielt Informationen darüber, welche Kategorien pro Teilkompetenz und Aufgabe von den Lernenden genannt wurden sowie welche Niveaus insgesamt und welches höchste Niveau von einer Person erreicht wurden. Falls Kategorien in den Antworten der Lernenden mehrfach angeführt wurden, zählten diese nur einfach.

Deskriptive Analysen

Die Ergebnisse der inhaltlichen Überprüfung der theoretischen Struktur des Kompetenzmodells (Grünkorn, Upmeier zu Belzen et al., 2014) und der Beschreibung von Besonderheiten in den Antwortstrukturen der Lernenden (vgl. S. 104ff. dieser Arbeit) basieren auf Berechnungen von Häufigkeiten. In der Publikation von Grünkorn, Upmeier zu Belzen et al. (2014) wird für die einzelnen Kategorien über prozentuale Häufigkeiten berichtet. Diese Häufigkeiten beziehen sich auf die Anzahl derjenigen Lernenden pro Teilkompetenz, die Aufgaben in der jeweiligen Teilkompetenz bearbeiteten. Die Summe aller prozentualen Häufigkeiten der Kategorien einer Teilkompetenz übersteigt bei einigen Teilkompetenzen 100 Prozent, weil Lernende in ihren Antworten mehrere Kategorien ansprachen. Bei anderen Teilkompetenzen lag die Summe aller prozentualen Häufigkeiten der Kategorien innerhalb einer Teilkompetenz unter 100 Prozent. Dies liegt daran, dass Schülerinnen und Schüler entweder nicht antworteten oder dass die Antworten nicht mit den Fragen korrespondierten.

Über diese prozentualen Häufigkeiten hinaus wurden die prozentuale Verteilung der Schülerinnen und Schüler über die drei Niveaus berechnet. Die Verteilung bezieht sich ebenfalls auf die Anzahl der Lernenden, die sich mit dieser Teilkompetenz beschäftigten. Für die Berechnung wurde nicht das höchste Niveau gewertet, sondern es wurden alle erreichten Niveaus eines Lernenden einbezogen. Somit können Aussagen darüber getroffen werden, welche Niveaus insgesamt bereits von den Lernenden erreicht wurden (vgl. Grünkorn, Upmeier zu Belzen et al., 2014).

Auch für die Beschreibung von Besonderheiten in den Antwortstrukturen (vgl. S. 104ff. dieser Arbeit) wurden Häufigkeiten für die identifizierten Kategorien

berechnet. Diese beziehen sich einerseits auf die Anzahl der Lernenden-Antworten pro Teilkompetenz und andererseits auf die Gesamtanzahl der Lernenden-Antworten über alle Teilkompetenzen hinweg.

Methoden der Item-Response-Theorie (IRT)

Um zu prüfen, ob die Aufgaben des entwickelten Leistungstests messmodell- bzw. raschkonform sind (vgl. S. 111ff. dieser Arbeit) und damit in Studien eingesetzt werden können, die einen probabilistischen Ansatz verfolgen (vgl. Kauertz, 2014; Neumann, 2014), wurden entsprechende Methoden der probabilistischen Testtheorie, der IRT (Moosbrugger, 2012), verwendet. In der IRT wird die Bearbeitung von Aufgaben als Indikator von latenten Fähigkeiten verstanden, die nicht direkt messbar sind. Die klassische Testtheorie (KTT) dagegen nimmt an, dass die erhobenen Testwerte unter Berücksichtigung von Messfehlern direkt auf die Fähigkeiten von Personen zurückzuführen sind. Somit hängt in der KTT die Personenfähigkeit von der Schwierigkeit der ausgewählten Items und der Stichprobe ab. Dies ist bei der IRT nicht der Fall, weil die Aufgabenschwierigkeiten ohne Berücksichtigung der Personenparameter ermittelt werden (*Stichprobenunabhängigkeit*, Moosbrugger, 2012). Die Personenparameter werden separat mit entsprechenden Methoden (z. B. der Weighted Maximum Likelihood oder der Plausible Values) geschätzt (Rost, 2004). In einem darauffolgenden Schritt werden in der IRT beide Parameter auf der gleichen Logit-Skala (*joint scale*) zusammengeführt (Rauch & Hartig, 2012; Moosbrugger, 2012). Die Ausprägungen der Aufgaben- und Personenparameter werden in Form von Logits angegeben. Diese Logits sind Wettquotienten, die die Wahrscheinlichkeit (Odds-Ratio) beschreiben, mit der eine Person ein Item richtig lösen wird. Sie werden aus der Wahrscheinlichkeit, ein Item richtig zu lösen, und der Gegenwahrscheinlichkeit berechnet (Rauch & Hartig, 2012; Rost, 2004). Je nach Schwierigkeit der Aufgabe ist diese weiter oben (schwierig) oder weiter unten (leicht) auf der Logit-Skala angeordnet. Analog dazu sind auch die Personenfähigkeiten angeordnet (unterer Bereich \triangleq geringe Fähigkeit; oberer Bereich \triangleq hohe Fähigkeit). Ist der Aufgabenparameter genauso groß wie der Personenparameter, so ist der Logit gleich Null und die Lösungswahrscheinlichkeit beträgt 50 Prozent. Theoretisch können die Parameter auf der Logit-Skala Werte zwischen plus und minus unendlich annehmen. Typi-

scherweise ergeben sich aber Werte im Bereich minus drei bis plus drei Logits (Rauch & Hartig, 2012).

Ausgehend von der Logit-Skala können Aussagen darüber getroffen werden, welche Teilkompetenzen oder Ausprägungen der jeweiligen Teilkompetenz eines Konstrukts schwieriger sind als andere (Kauertz, 2014). Ein probabilistischer Ansatz ist auch aus mehreren anderen Gründen vorteilhaft: In einem solchen Ansatz können die Reaktionen von Personen auf raschskalierte Aufgaben vorhergesagt werden, auch wenn diese Aufgaben von den Personen gar nicht bearbeitet wurden. Dies ist deshalb möglich, weil bei der Bestimmung der Personenfähigkeit nur wichtig ist, wie viele Aufgaben eine Person gelöst hat, und nicht, welche Aufgabe sie bearbeitet hat (*spezifische Objektivität*, Rasch, 1977; zitiert nach Neumann, 2014, S. 358). Ähnliches gilt für die Bestimmung der Aufgabenschwierigkeit, die von der Anzahl der Personen abhängt, die sie gelöst haben. Unbeachtet bleibt die Information, wer sie gelöst hat. Dadurch erlaubt die IRT auch die Auswertung von Datensätzen mit einem MMX-Design. Mit Methoden der KTT sind solche Datensätze nur stark eingeschränkt auswertbar, weil für weitere KTT-Analysen die fehlenden Datensätze zunächst mit Methoden der IRT geschätzt werden müssen (Prenzel, Walter & Frey, 2007).

Die fehlenden Werte werden durch Maximum-Likelihood-Schätzungen (z. B. marginal Maximum-Likelihood, conditional Maximum-Likelihood) ersetzt. Das bedeutet, dass aus der gemeinsamen Stichprobe von Personen und Aufgaben die jeweiligen Parameter so geschätzt werden, dass die Wahrscheinlichkeit, mit der die empirischen Daten **zu dem Modell „passen“, maximiert wird** (Bühner, 2010; Rost, 2004). Für diese Schätzung ist je nach Dimensionalität des Modells ein spezifischer Schätzalgorithmus zu verwenden (Wu, Adams & Wilson, 2007). Die Schätzverfahren sind durch einen iterativen Prozess gekennzeichnet, das heißt, es wird so lange iterativ geschätzt, bis die bestmögliche Lösung erreicht ist. Für die Schätzung spielen sogenannte Knoten (*nodes*) eine entscheidende Rolle. Sie kennzeichnen die Stellen in der Verteilung, an denen eine Schätzung vorgenommen wird. Je mehr *nodes* bestimmt werden, desto näher kommt die Schätzung der eigentlichen Verteilung bzw. desto genauer wird geschätzt (Wu et al., 2007). Für die Schätzung der Personenparameter stehen verschiedene Methoden zur Verfügung (Rost, 2004). Auf diese verschiedenen

Schätzverfahren wird aber nicht näher eingegangen, weil in dieser Forschungsarbeit die Überprüfung der Aufgaben im Vordergrund steht.

Vorbereitung und Durchführung der IRT-Analysen

In der IRT stehen verschiedene Modelle zur Verfügung, mit denen sich Personen- und Aufgabenparameter bestimmen lassen. Das in dieser Arbeit verwendete Modell muss einerseits eine mehrdimensionale Skalierung und andererseits mehrstufige Antwortkategorien zulassen. Daher fand das ordinale, mehrdimensionale Rasch-Modell – das Partial-Credit-Modell – Verwendung (Masters, 1982).

Die Datenmatrix für die Analysen mithilfe des Partial-Credit-Modells umfasste Informationen darüber, welches höchste Niveau für jede Aufgabe erreicht wurde. Für jede Aufgabe waren vier Antwortkategorien möglich, die für Folgendes stehen:

- 0: unbeantwortete Aufgabe, nicht mit der Frage korrespondierende Antworten und Antworten auf dem basalen Niveau
- 1: Antworten auf dem Niveau I des Kompetenzmodells
- 2: Antworten auf dem Niveau II des Kompetenzmodells
- 3: Antworten auf dem Niveau III des Kompetenzmodells

Wie beschrieben, sind unter der Antwortkategorie 0 verschiedene Kategorien zusammengefasst, die zunächst getrennt erfasst und notiert wurden. Unter der Antwortkategorie 0 sind Personen subsumiert, deren Kompetenzen nicht denen auf Niveau I, II oder III der jeweiligen Teilkompetenz entsprechen. Unbeantwortete Aufgaben, das heißt Antworten von Personen, welche die Aufgaben ausließen oder übersprangen, wurden in dieser Studie auf die gleiche Weise interpretiert. Diese Interpretation ist deshalb möglich, weil die Lernenden in der Testung die Aufgaben im offenen Format zuerst bearbeiteten und ihnen ausreichend Zeit für die Bearbeitung zur Verfügung gestellt wurde. Eine Rekodierung der fehlenden Daten ist unter diesen Umständen nach Rost (2004, S. 87) möglich.

Die Auswertung erfolgte mit dem Programm ACER ConQuest Version 3.0.1 (Wu, Adams & Wilson, 2012). Das Programm nutzt die marginale Maximum-Likelihood-Methode für die Parameterschätzung (Wu et al., 2007). Bei dieser Methode werden die Personen- und Aufgabenparameter nicht simultan ge-

schätzt. Darüber hinaus liegt dieser Methode zugrunde, dass die Personenparameter normalverteilt sind und die Stichprobe zufällig gezogen wurde (Bühner, 2010). Mit Blick auf die Schätzalgorithmen fanden zwei verschiedene Algorithmen Verwendung (Wu et al., 2007): die Gauss-Hermite-Quadrature nach Bock und Aitken (1981; zitiert nach Wu et al., 2007) und die Monte-Carlo-Methode nach Volodin und Adams (1995; zitiert nach Wu et al., 2007). Die Verwendung von zwei Algorithmen wird damit begründet, dass die Aufgaben für drei Strukturmodelle und damit für drei Hypothesen über die Struktur von Modellkompetenz geprüft wurden: Beim **eindimensionalen Strukturmodell** lassen sich keine Dimensionen unterscheiden und alle Aufgaben laden auf den globalen Faktor Modellkompetenz. Das **zweidimensionale Strukturmodell** basiert auf der Annahme, dass Modellkompetenz in die zwei Dimensionen **Kenntnisse über Modelle** und **Modellbildung** unterteilt werden kann (vgl. Upmeyer zu Belzen & Krüger, 2010). Im Gegensatz dazu beruht das **fünfdimensionale Strukturmodell** auf der Annahme, dass sich hinsichtlich der Modellkompetenz fünf Teilkompetenzen (**Eigenschaften von Modellen, Alternative Modelle, Zweck, Testen** und **Ändern von Modellen**) unterscheiden lassen (vgl. Upmeyer zu Belzen & Krüger, 2010). Vor diesem theoretischen Hintergrund und vor dem Hintergrund, dass Studien zur Dimensionalität des Kompetenzmodells zu unterschiedlichen Ergebnissen gekommen sind (Krell & Krüger, 2011; Krell, 2013; Krell et al., 2012a; Terzer, 2013), ist die Prüfung des Leistungstests jeweils für das ein-, zwei und fünfdimensionale Strukturmodell vorzunehmen. Um eine möglichst genaue Modellschätzung für das jeweilige Strukturmodell zu erreichen, wurde in ConQuest eine Vorgehensweise mit zwei Analysedurchgängen gewählt. Zunächst wurden in einem ersten Durchlauf Initial-Parameter geschätzt, die dann für einen zweiten Analysedurchgang als Startwerte eingesetzt wurden. Für die Überprüfung der Aufgaben bedeutete dies, dass auf der Basis der Gauss-Hermite-Quadrature nach Bock und Aitken (1981; zitiert nach Wu et al., 2007) in einem ein- und zweidimensionalen Strukturmodell die Initial-Parameter zunächst mit 10 **nodes** geschätzt wurden. Die geschätzten Parameter wurden dann für einen zweiten Durchlauf mit 20 **nodes** verwendet. Für das fünfdimensionale Strukturmodell wurde die Monte-Carlo-Methode nach Volodin und Adams (1995; zitiert nach Wu et al., 2007) mit 400 **nodes** im ersten und 5 000 **nodes** im zweiten Durchlauf genutzt. Weiterhin wurden bei allen Analysen die Mittelwerte der Personenfähigkeiten auf Null restringiert, um In-

formationen über die Aufgaben zu gewinnen (*fallzentrierte Analyse*, Moosbrugger, 2012).

Überprüfung des Leistungstests mit Methoden der IRT

Zur Überprüfung des Leistungstests mit den beschriebenen Schätzverfahren wurden ähnlich wie bei Neumann (2014) zunächst die Modellanpassung beurteilt und dann die Aufgabenparameter und die Personenfähigkeiten geprüft.

Modellanpassung

Ein erster Schritt in der Bewertung der Analysen mithilfe der oben beschriebenen Verfahren ist die Beurteilung, inwiefern die Anpassung des Modells in den jeweiligen Schätzverfahren erfolgreich war. Daher ist zu ermitteln, ob die Schätzungen konvergiert sind und damit eine optimale Lösung gefunden wurde. Ist dies nicht der Fall, können keine weiteren Analysen erfolgen (Neumann, 2014).

Aufgabenparameter

Nachdem geprüft wurde, inwiefern das Modell zu verlässlichen Ergebnissen kommt, werden mehrere Kriterien zur Beurteilung der Eignung des Tests für weitere Interpretationsschritte herangezogen (Kauertz, 2014). Neben klassischen Gütekriterien wurden hier IRT-basierte Merkmale sowohl auf Aufgabenebene als auch auf der Ebene der Antwortkategorien betrachtet. Tabelle 6 zeigt eine Übersicht über die zur Beurteilung der Aufgaben herangezogenen Kennwerte.

Tab. 6: Kennwerte zur Beurteilung des Leistungstests in einem probabilistischen Ansatz (Bühner, 2010; Hartig, 2009; Kelava & Moosbrugger, 2012)

Ebene der Aufgabe	Ebene der Antwortkategorien
<ul style="list-style-type: none">- Trennschärfe- Schwierigkeit- Aufgaben-Fit (wMNSQ und t-Werte)	<ul style="list-style-type: none">- Punktbiseriale Korrelationen- Thurstonian Thresholds- Item-Deltas- Aufgaben-Fit (wMNSQ und t-Werte)- Häufigkeiten

Auf der Aufgabenebene wurde zunächst die *Trennschärfe* bestimmt. ConQuest Version 3.0.1 gibt hierfür sogenannte Item-Rest Correlations pro Aufgabe aus. Für die Berechnung der Item-Rest-Correlation wird eine Part-whole-Korrektur

vorgenommen, bei der die betreffende Aufgabe nicht in den Skalenwert eingeht. Auf diese Weise kann bei einer geringen Aufgabenanzahl verhindert werden, dass die Trennschärfe überschätzt wird (Bühner, 2010). Die Trennschärfe gibt an, wie gut eine Aufgabe eine Skala widerspiegelt. Die Skala bezieht sich dabei auf Aufgaben, die inhaltlich dasselbe messen. Aus diesem Grund wurden die Trennschärfen der Aufgaben ausschließlich pro Teilkompetenz berechnet (Bühner, 2010; Kelava & Moosbrugger, 2012). Als gute Trennschärfen gelten part-whole-korrigierte Trennschärfe-Werte von .4 bis .7 (Kelava & Moosbrugger, 2012).

Auf der Ebene der Antwortkategorien sind für die Trennschärfe die **punktbiserialen Korrelationen** zu ermitteln. Sie beschreiben klassisch die Korrelationen zwischen dem Vorliegen einzelner Antwortkategorien und der Gesamtleistung im Test. Die Werte sollten möglichst von negativen Werten für die niedrigen Kategorien zu positiven für die höheren Kategorien ansteigen (Bühner, 2010).

Als weiteren Kennwert generiert ConQuest **Aufgabenschwierigkeiten** (Estimate). Diese berechnen sich aus den Mittelwerten der Thurstonian-Threshold-Parameter für die Antwortkategorien. Die **Thurstonian Thresholds** (τ) geben diejenige Stelle auf der Skala an, an der die Wahrscheinlichkeit, dass die Antwortkategorie k oder eine höhere erzielt wird, 50 Prozent beträgt. $\tau < 0$ kennzeichnen Aufgaben, die für die Stichprobe generell leichter sind als $\tau > 0$. Die Thurstonian Thresholds lassen keine Vertauschungen bzw. ungeordneten Schwellen zu, sodass sie nicht den Schwellenparametern (Item-Deltas) des Modells entsprechen (Wu et al., 2007).

Um zu prüfen, ob die Aufgaben raschkonform sind, wurden **Aufgaben-Fit-Maße** herangezogen (Bond & Fox, 2007). ConQuest berechnet hierzu für jede Aufgabe und für jede Antwortkategorie gewichtete (weighted) Mean-Square-(wMNSQ)-Fit-Statistiken. Sie beschreiben die Abweichung der Aufgabe bzw. der Antwortkategorie von einem festgelegten Erwartungswert. Auf diese Weise kann überprüft werden, wie gut die Aufgabe zu dem Modell passt. Werte in einem Bereich von $0.8 < wMNSQ < 1.2$ sind akzeptabel (Bond & Fox, 2007). In Verbindung hiermit sind die zugehörigen t-Werte zu betrachten. Diese geben Auskunft darüber, ob die Abweichungen vom Erwartungswert signifikant sind. t-Werte > 2 bzw. < -2 kennzeichnen einen signifikant zu guten Fit bzw. einen signifikant zu schlechten Fit (Neumann, 2014).

Bei der Verwendung des Partial-Credit-Modells ist außerdem zu prüfen, ob geordnete Schwellen (δ , Item-Deltas) vorliegen. **Item-Deltas** kennzeichnen die Stelle, an der sich benachbarte Kategorienfunktionen überschneiden. An dieser Stelle ist die Lösungswahrscheinlichkeit für beide Antwortkategorien gleich groß. Sind die Schwellenparameter geordnet, so ist davon auszugehen, dass niedrigere Antwortkategorien auch mit niedrigeren Fähigkeiten einhergehen (für ein Beispiel aus dem hier beschriebenen Test vgl. Abb. 10, S. 114). Liegen die Schwellen ungeordnet vor, ist anzunehmen, dass die aufeinander folgenden Antwortkategorien nicht den aufeinander folgenden Abschnitten der Personenfähigkeiten entsprechen (für ein Beispiel aus dem hier beschriebenen Test vgl. Abb. 11, S. 114). Die betreffende Antwortkategorie erhält auf dem **latenten Kontinuum keinen „eigenen“ Abschnitt; das heißt, es gibt keinen Abschnitt, an dem die Wahrscheinlichkeit, in dieser Kategorie zu antworten, am höchsten ist** (Rost, 2004, S. 205).

Alle genannten Kennwerte hängen insbesondere von den **Häufigkeiten** der Antwortkategorien ab. Aus diesem Grund wurden diese ebenfalls betrachtet. Antwortkategorien, deren Häufigkeit geringer als fünf Prozent oder zehn Prozent ist, können zu nicht akzeptablen Werten führen (Hartig, 2009).

Personenfähigkeiten

Neben den Aufgabenparametern wurden zudem die Personenfähigkeiten untersucht. ConQuest generiert hierzu pro Aufgabe und Antwortkategorie die Verteilung der Personenparameter (Wu et al., 2007). Bei den Antwortkategorien handelt es sich um ordinale Kategorien. Daher sollten die Mittelwerte für die Personen mit den einzelnen Antwortkategorien möglichst kontinuierlich steigen (Hartig, 2009).

Zur Beurteilung des Tests sind außerdem unter anderem die Reliabilität und die Varianz zu betrachten. Für die Einschätzung der Messgenauigkeit von Tests auf der Basis von MMX-Designs wurde im Rahmen der Analysen in ConQuest nach der Expected-a-posteriori-Methode (EAP) die Zuverlässigkeit (EAP/PV-Reliabilität) des Instruments bestimmt (Rost, 2004). Diese Werte sind auf gleiche Weise wie Cronbachs Alpha zu interpretieren ($r > .8 \triangleq$ gut, Rost, 2004). Allerdings ist zu beachten, dass für Tests auf Populationsebene auch geringere Reliabilitätswerte akzeptabel sind (Klieme, Avenarius et al., 2007; Leutner et al., 2007; Terzer et al., 2013). Zudem steht die Reliabilität in enger Verbin-

derung mit der Varianz und der Testlänge (Amelang & Zielinski, 2002). Mit Blick auf die Varianz (Var) der Personenverteilung ist eine $Var > 1$ anzustreben (Neumann, 2014; Rost, 2004). $Var < 1$ könnte auf einen Test hinweisen, der nicht deutlich zwischen den Personen unterscheidet und damit das Fähigkeitsspektrum nicht ausreichend abdeckt (Neumann, 2014, S. 365). Um Letzteres zu überprüfen, sind Wright Maps heranzuziehen, die bei mehrstufigen Aufgaben den Aufgabenparameter *Thurstonian Thresholds* (rechts) und die Personenfähigkeiten (links) auf einer Logit-Skala grafisch darstellen (für ein Beispiel aus dem hier beschriebenen Test vgl. Abb. 12, S. 116).

5.2 Ergebnisse der Analyse von Modellkompetenz

Basierend auf dem beschriebenen methodischen Vorgehen, sind die Ergebnisse der inhaltlichen Überprüfung des Kompetenzmodells der Modellkompetenz in der folgenden Publikation veröffentlicht und in Kapitel 5.2.1 (vgl. S. 100ff. dieser Arbeit) zusammengefasst:

- Grünkorn, J., Upmeier zu Belzen, A. & Krüger, D. (2014). Assessing Students' Understandings of Biological Models and Their Use in Science to Evaluate a Theoretical Framework. *International Journal of Science Education*, 1-34. doi: 10.1080/09500693.2013.873155

Die Ergebnisse der Beschreibung von Besonderheiten in den Antwortstrukturen der Lernenden (vgl. S. 104ff. dieser Arbeit) und der Analyse im Rahmen der IRT (vgl. S. 111ff. dieser Arbeit) sind bisher unveröffentlicht und werden daher in der vorliegenden Arbeit umfassend beschrieben.

5.2.1 Beitrag 3: Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework

Grünkorn, J., Upmeier zu Belzen, A. & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education*, 1-34. doi: 10.1080/09500693.2013.873155

Zusammenfassung

Dieser Beitrag beschreibt die inhaltliche Überprüfung der theoretischen Struktur des Kompetenzmodells der Modellkompetenz. Die zentrale Fragestellung war:

- F1 Inwieweit werden mit der theoretischen Struktur des Kompetenzmodells der Modellkompetenz die Perspektiven von Lernenden auf Modelle und auf die Modellbildung umfassend und adäquat beschrieben?

Diese Fragestellung lässt sich in die folgenden Detailfragen untergliedern:

- F1.1 Welche Perspektiven (Kategorien) von Lernenden auf Modelle und auf die Modellbildung lassen sich für die drei Niveaus der jeweiligen Teilkompetenz des Kompetenzmodells der Modellkompetenz identifizieren?
- F1.2 Welche zusätzlichen Perspektiven (Kategorien) von Lernenden auf Modelle und auf die Modellbildung lassen sich zu den jeweiligen Teilkompetenzen des Kompetenzmodells der Modellkompetenz identifizieren?
- F1.3 In welchen Häufigkeiten treten die Perspektiven (Kategorien) von Lernenden auf Modelle und auf die Modellbildung in der untersuchten Stichprobe auf?

Das Kompetenzmodell von Upmeier zu Belzen und Krüger (2010) basiert unter anderem auf empirischen Studien von Crawford und Cullin (2005), Grosslight et al. (1991) sowie Justi und Gilbert (2003). Daher wird erwartet, dass die theoretische Struktur des Kompetenzmodells der Modellkompetenz die Perspektiven von Lernenden auf Modelle und auf die Modellbildung hinreichend beschreibt (H1). Im Hinblick auf die Häufigkeiten kann angenommen werden, dass Kategorien der Niveaus I und II häufiger auftreten als Kategorien des Niveaus III (H2). Diese Annahme lässt sich daraus ableiten, dass Vorgängerstudien zu Modellen und zur Modellbildung festgestellt haben, dass Lernende eher eine mediale Perspektive auf Modelle aufweisen (z. B. Grosslight et al., 1991; Grünkorn & Krüger, 2012; Grünkorn et al., 2011; Krell, Upmeier zu Belzen et al. 2012b; Schwarz et al., 2009; Schwarz & White, 2005; Terzer, 2013; Treagust et al., 2002; Trier & Upmeier zu Belzen, 2009).

Auf der Grundlage von 3 531 Antworten der Lernenden konnten insgesamt 42 Kategorien identifiziert werden, von denen 29 Kategorien die Niveaus der jeweiligen Teilkompetenzen ausdifferenzieren. Tabelle 7 zeigt einen Überblick über die identifizierten Kategorien (F1.1) und deren Häufigkeiten (F1.3) innerhalb der fünf Teilkompetenzen. Zudem präsentiert die Tabelle die prozentualen Verteilungen der von den Lernenden erreichten Niveaus innerhalb der jeweiligen Teilkompetenzen (F1.3).

Tab. 7: Kategorien und ihre Häufigkeiten (%)^a innerhalb der Niveaus sowie Gesamthäufigkeiten (%)^b pro Niveau der Teilkompetenz (verändert nach Grünkorn, Upmeyer zu Belzen & Krüger, 2014)

Teilkompetenzen	Komplexität		
	Niveau I	Niveau II	Niveau III
Eigenschaften von Modellen ($n_{Lernende} = 692$)	Modell als Kopie (27 %)	Modell ist in Teilen eine Kopie (9 %)	Modell als hypothetische Darstellung (4 %)
	Modell mit großer Ähnlichkeit (36 %)	Modell als eine mögliche Variante (7 %)	
	Modell entspricht (nicht) subjektiver Vorstellung vom Original (6 %)	Modell als fokussierte Darstellung (0.3 %)	
Gesamt	69 %	17 %	4 %
Alternative Modelle ($n_{Lernende} = 705$)	Unterschiedliche Modellobjekteigenschaften (44 %)	Unterschiedliche inhaltliche Schwerpunkte (31 %)	Unterschiedliche Annahmen (8 %)
			Unterschiedliche Annahmen mit Anwendungsperspektive (1 %)
Gesamt	44 %	31 %	8 %
Zweck von Modellen ($n_{Lernende} = 706$)	Modell zum Darstellen eines Sachverhalts (52 %)	Modell zum Erkennen von Zusammenhängen (45 %)	Modell zum Überprüfen von abstrakten Ideen (19 %)
		Modell zum Erklären von Zusammenhängen (6 %)	Modell zum Überprüfen von konkreten Ideen (5 %)
Gesamt	52 %	50 %	24 %
Testen von Modellen ($n_{Lernende} = 711$)	Überprüfung des Materials (6 %)	Vergleich von Original und Modell (33 %)	Überprüfung von Hypothesen (6 %)
	Überprüfung der Grundvoraussetzungen (28 %)	Vergleich und Passung von Original und Modell (34 %)	Überprüfung von Hypothesen mit Forschungsdesign (2 %)
Gesamt	30 %	68 %	8 %
Ändern von Modellen ($n_{Lernende} = 712$)	Ändern zur Verbesserung des Modellobjekts (22 %)	Ändern bei mangelnder Passung mit dem Original (61 %)	Ändern bei Erkenntnissen aus Modellexperiment (1 %)
	Ändern bei Fehlern im Modellobjekt (5 %)	Ändern bei neuen Erkenntnissen über das Original (4 %)	
	Ändern bei Nichterfüllung der Grundvoraussetzungen (9 %)	Ändern bei Veränderungen des Originals (5 %)	
Gesamt	31 %	68 %	1 %

Anmerkungen. ^a Die Häufigkeiten der Kategorien in Prozent wurden auf der Basis der Anzahl der Lernenden berechnet, die Antworten in der jeweiligen Teilkompetenz generiert haben.

^b Die prozentualen Angaben für die gesamte Teilkompetenz beziehen sich auf alle erreichten Niveaus der Lernenden und nicht auf das höchste Niveau.

In Bezug auf die zweite Fragestellung (F1.2) konnten zusätzliche Kategorien für die Teilkompetenzen *Alternative Modelle*, *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen* identifiziert werden. Tabelle 8 skizziert zusätzlich generierte Kategorien (F1.2) und ihre Häufigkeiten (F1.3). In den Analysen sind diese zusätzlichen Kategorien unter *weitere Kategorien* und *basale Kategorien* subsumiert. Erstere beschreiben alltägliche Perspektiven auf Modelle, die mit Blick auf die Nutzung von Modellen im erkenntniswissenschaftlichen Prozess eine untergeordnete Rolle spielen. Die basalen Kategorien, die in den Teilkompetenzen *Alternative Modelle*, *Testen von Modellen* und *Ändern von Modellen* identifiziert wurden, umfassen Perspektiven, die als basale Niveaus der jeweiligen Teilkompetenzen zu diskutieren sind.

Tab. 8: Zusätzliche Kategorien und ihre Häufigkeiten (%)^{*} zu den Teilkompetenzen (verändert nach Grünkorn, Upmeier zu Belzen & Krüger, 2014)

Teilkompetenzen	Zusätzliche Kategorien	
	Weitere Kategorien ^a	Basale Kategorien ^b
Eigenschaften von Modellen ($n_{Lernende} = 692$)	-	-
Alternative Modelle ($n_{Lernende} = 705$)	-	Modelle sind gleich (4 %) Verschiedene Modelle zu unterschiedlichen Originalen (13 %) Nur ein endgültiges und richtiges Modell (1 %)
Zweck von Modellen ($n_{Lernende} = 706$)	Modell als Spielzeug (0.4 %) Modell zu Dekorationszwecken (2 %) Modell als Ersatzobjekt (1 %) Modell als Bauplan (0.1 %) Modell zu Werbezwecken (1 %)	-
Testen von Modellen ($n_{Lernende} = 711$)	-	Keine Testung des Modells (1 %)
Ändern von Modellen ($n_{Lernende} = 712$)	Ändern des Modellzustandes (2 %)	Kein Anlass für ein Änderung (3 %) Ändern zur Darstellung unterschiedlicher Originale (7 %)

Anmerkungen. * Die Häufigkeiten in Prozent wurden auf der Basis der Anzahl der Lernenden berechnet, die Antworten in der jeweiligen Teilkompetenz generiert haben.

^a Weitere Kategorien beschreiben alltägliche Perspektiven auf Modelle.

^b Basale Kategorien beschreiben Perspektiven, die als basale Niveaus für die jeweilige Teilkompetenz zu diskutieren sind.

5.2.2 Beschreibung von Besonderheiten in den Antwortstrukturen der Lernenden

Um ein vertieftes Verständnis der Sichtweisen der Lernenden auf Modelle zu bekommen, ist es sinnvoll, Besonderheiten in den Antwortstrukturen von Lernenden zu analysieren. Eine relevante Rolle scheinen aufgrund von Hinweisen aus Untersuchungen didaktisch orientierte Perspektiven zu spielen, die sich auf die Kommunikation und die Zugänglichmachung von Phänomenen beziehen.

So wiesen Grosslight et al. (1991) in ihren Kategorienbeschreibungen darauf hin, dass Schülerinnen und Schülern den Kommunikations- und Zugänglichkeitsaspekt von Modellen wahrnahmen. Ingham und Gilbert (1991) beschrieben in ihrer Untersuchung mit Studierenden der Chemie die Perspektiven *the use of models for personal purposes* und *the use of models for the purposes of others* und stuften diese als relevant für die Entwicklung eines elaborierten Modellverständnisses ein. Auch Schwarz et al. (2009) nehmen an, dass insbesondere der Kommunikationsaspekt bei Lernenden aber auch bei Forschenden wichtig ist, um Originale besser verstehen zu können (vgl. S. 27ff. dieser Arbeit). Aus diesen Gründen stehen bei der Analyse vor allem die didaktisch orientierten Perspektiven mit Blick auf die Kommunikation und die Zugänglichmachung von Phänomenen im Mittelpunkt. Mit Bezug darauf soll die folgende Fragestellung in dieser Forschungsarbeit untersucht werden:

F2 Inwiefern lassen sich didaktisch orientierte Perspektiven mit Blick auf die Kommunikation und die Zugänglichmachung bei den Antworten der Schülerinnen und Schüler in den fünf Teilkompetenzen beschreiben?

Aufbauend auf den Hinweisen von Grosslight et al. (1991) und den Befunden von Ingham und Gilbert (1991) für Studierende der Chemie ist anzunehmen, dass sich auch bei den Antworten der Schülerinnen und Schülern der gewählten Stichprobe in den fünf Teilkompetenzen des Kompetenzmodells didaktisch orientierte Perspektiven beschreiben lassen (H3).

Ingham und Gilbert (1991) zeigten, dass diese didaktisch orientierten Perspektiven bei Studierenden in Verbindung mit unterschiedlichen Ausprägungen eines wissenschaftstheoretischen Modellverständnisses auftreten. Können auch bei Schülerinnen und Schülern didaktisch orientierte Perspektiven identifiziert werden und treten diese in unterschiedlichen Teilkompetenzen und in Verbin-

dung mit unterschiedlichen Niveaus (mediale und methodische Perspektiven) auf, könnten in weiteren Studien Zusammenhänge zwischen den Teilkompetenzen und den didaktisch orientierte Perspektiven untersucht werden. Stellen sich solche Zusammenhänge dar, könnte hier ein zentraler Ansatzpunkt für eine Förderung von Modellkompetenz liegen. Es lässt sich vermuten, dass die Weiterentwicklung didaktisch orientierter Perspektiven, die alle oder zumindest mehrere Teilkompetenzen und Niveaus berühren, auch eine Weiterentwicklung dieser Teilkompetenzen nach sich zieht. Vor diesem Hintergrund ist die folgende Fragestellung zu untersuchen:

F3 In Verbindung mit welcher medialen und methodischen Perspektive (bzw. Niveaus) treten didaktisch orientierte Perspektiven mit Blick auf die Kommunikation und die Zugänglichmachung bei den Antworten der Schülerinnen und Schülern in den Teilkompetenzen auf?

Ausgehend von den Befunden von Ingham und Gilbert (1991), lässt sich vermuten, dass von Schülerinnen und Schülern didaktisch orientierte Perspektiven sowohl mit einer medialen als auch mit einer methodischen Perspektive auf Modelle beschrieben werden (H4).

Neben der Fokussierung auf diese Perspektiven sollen auf explorative Weise weitere Besonderheiten in den Antwortstrukturen von Lernenden identifiziert werden. Auf diese Weise können mögliche Erklärungsansätze für die Perspektiven von Schülerinnen und Schülern auf Modelle und auf den Modellbildungsprozess geliefert werden.

Analyse der didaktisch orientierten Perspektiven

In den insgesamt 3 531 Antworten der Lernenden, die zu den fünf Teilkompetenzen des Kompetenzmodells der Modellkompetenz gegeben wurden, ließen sich die Kategorien *Modell als Mittel der Zugänglichkeit* und *Modelle als Mittel der Verständlichkeit/Kommunizierbarkeit* identifizieren (F2.1). Tabelle 9 beschreibt beide Kategorien, präsentiert Beispielaussagen von Schülerinnen und Schülern und nennt prozentuale Häufigkeiten.

Tab. 9: Beschreibung der didaktisch orientierten Perspektiven mit Beispielaussagen und Häufigkeiten (%)* bezogen auf alle Teilkompetenzen

Name und Beschreibung der Kategorie	Beispielaussagen von Lernenden (ID des Lernenden)	Häufigkeit (%)
<p>Modell als Mittel der Zugänglichkeit</p> <ul style="list-style-type: none"> - dient dazu, Dinge für einen bestimmten Adressaten zugänglich bzw. direkt erfahrbar zu machen 	<p>„Es hat den Zweck, Dinge darzustellen, die nicht für alle Menschen zugänglich sind“ (Fd794)</p> <p>„Modelle können für Leute erfunden werden, die so etwas im realen Leben nicht kennen, und durch Modelle können sie so ungefähr sehen, wie das aussieht“ (Fd575)</p>	5
<p>Modell als Mittel der Verständlichkeit/ Kommunizierbarkeit</p> <ul style="list-style-type: none"> - verbessert das Verständnis und die Kommunizierbarkeit, erleichtert die Vermittlung 	<p>„Da es schon in der 4. Klasse Bio gibt, aber die Kinder noch nicht die komplexe Darstellung (A) verstehen, gibt es Modell B und C, um es auch kleineren Kindern veranschaulichen zu können [...]“ (Fb142)</p> <p>„Das Modell wird verändert, damit es anderen erleichtert, den Ablauf des Mundes zu verstehen“ (Fb438)</p>	12

Anmerkung. * Die Häufigkeiten in Prozent beziehen sich auf 3 531 Antworten der Lernenden, die in den fünf Teilkompetenzen des Kompetenzmodells auftraten.

Überdies ließen sich beide Kategorien in allen Teilkompetenzen des Kompetenzmodells feststellen (F2.1). Tabelle 10 gibt einen Überblick über die Häufigkeiten pro Teilkompetenz.

Tab. 10: Die Häufigkeiten (%)* der didaktisch orientierten Kategorien in den fünf Teilkompetenzen des Kompetenzmodells

Teilkompetenzen	Häufigkeiten (%)	
	Modell als Mittel der Zugänglichkeit	Modell als Mittel der Verständlichkeit/ Kommunizierbarkeit
Eigenschaften von Modellen ($n_{Lernende} = 692$)	9	4
Alternative Modelle ($n_{Lernende} = 705$)	6	26
Zweck von Modellen ($n_{Lernende} = 706$)	8	27
Testen von Modellen ($n_{Lernende} = 711$)	0.6	1
Ändern von Modellen ($n_{Lernende} = 712$)	1	2

Anmerkung. * Die Häufigkeiten in Prozent wurden auf der Basis der Anzahl der Lernenden berechnet, die Antworten in der jeweiligen Teilkompetenz generiert haben.

Im Hinblick auf den Zusammenhang zwischen einer wissenschaftstheoretischen Sicht auf Modelle (medial bzw. methodisch) und der Erwähnung der Kategorie *Modell als Mittel der Zugänglichkeit* konnten qualitativ Unterschiede

beschrieben werden (F2.2). Diese werden an einigen Beispielaussagen dargestellt: So antwortete beispielsweise Schüler 845 auf die Frage, inwieweit das Modell der Biomembran so aussehe wie eine Biomembran (Teilkompetenz **Eigenschaften von Modellen**):

„Die Biomembran kann man nicht direkt sehen, dafür braucht man dann ein Mikroskop und baut ein Modell. Ich weiß von meinem Vater, dass man schon sehr genau unter dem Mikroskop sehen kann, wie etwas aussieht. Eigentlich ist das Modell ja nur eine sehr starke Vergrößerung. Daher sieht das Modell schon so aus wie eine echte Biomembran.“

Schüler 845 nimmt das Modell als ein Hilfsmittel wahr, mit dem Personen nicht sichtbare bzw. nicht greifbare Phänomene in der Welt zugänglich werden. Damit beschreibt er die Sichtweise auf Modelle als Mittel der Zugänglichkeit. Zugleich verbindet Schüler 845 diese didaktisch orientierte Perspektive mit einer medialen Sicht auf die Eigenschaften von Modellen und einer naiv-realistischen Sichtweise auf die Welt, indem er Modelle als Vergrößerung bzw. als Kopie versteht (Niveau I).

Eine andere Beziehung zwischen einer wissenschaftstheoretischen Sicht auf Modelle (medial und methodisch) und der Erwähnung der Kategorie **Modell als Mittel der Zugänglichkeit** liegt bei Schülerin 293 vor. Sie äußerte sich zu der Frage, warum es zu dieser einer Biomembran drei verschiedene Modelle gebe (Teilkompetenz **Alternative Modelle**):

„[...] Da man die Biomembran nicht erkennen kann, weiß man auch nicht genau, wie sie aufgebaut ist. Deswegen muss man erst zu jeder Theorie verschiedene Modelle bauen, die man dann überprüft.“

Auch diese Schülerin nimmt das Modell als Mittel der Zugänglichkeit wahr und sieht es als Voraussetzung dafür an, ein Phänomen begreifen zu können. Im Gegensatz zum Schüler 845 verbindet diese Schülerin die didaktisch orientierte Perspektive mit einer methodischen Sichtweise auf alternative Modelle (Niveau III) und nimmt Modelle als Mediatoren wahr, die zwischen Theorie und Phänomen vermitteln.

Diese unterschiedlichen Beziehungen zeigten sich auch für die Kategorie **Modelle als Mittel der Verständlichkeit/Kommunizierbarkeit**. Schülerin 899 ant-

wortete auf die Frage, welchen Zweck das Darm-Modell erfülle (Teilkompetenz **Zweck von Modellen**):

„Es könnte dazu da sein, dass man in Schulen [...] sehen/zeigen kann, wie dick oder dünn der Darm ist. [...] Sie helfen einem, das Thema besser zu verstehen. Wenn man etwas bildlich sehen kann, ist es leichter, [et]was zu verstehen.“

Diese Schülerin beschreibt eine mediale Sicht auf Modelle als Mittel der Veranschaulichung (Niveau I), die sie mit dem didaktischen Zweck verbindet, ein besseres Verständnis des Phänomens zu bekommen. Hier liegt der Fokus darauf, Wissen auf einfache Weise vermittelt zu bekommen.

Schülerin 958 antwortete hingegen auf die Frage, warum es zu dieser einen Biomembran drei verschiedene Modelle gebe (Teilkompetenz **Alternative Modelle**):

„Verschiedene Wissenschaftler haben die Modelle gemacht, also haben sie auch verschiedene Vorstellungen zur Veranschaulichung. [...] Das dient der Diskussion und dem Austausch mit anderen über den Aufbau der Biomembran.“

Diese Schülerin nimmt die gezeigten Modelle als hypothetische Rekonstruktionen wahr (Niveau III), die verschiedene Vorstellungen repräsentieren. Zugleich hebt sie die Rolle der Modelle als Mittel der Kommunikation hervor, da sie zum Austausch und zur Diskussion herangezogen werden. Im Gegensatz zur vorherigen didaktisch orientierten Perspektive geht es hier weniger darum, Wissen zu vermitteln, als vielmehr darum, einen Austausch zu schaffen und damit das Phänomen besser zu verstehen. Daher beschreibt sie eine wissenschaftliche Sicht auf Modelle.

Analyse weiterer Besonderheiten in den Antwortstrukturen

Insgesamt ließen sich drei weitere Besonderheiten feststellen, die sich größtenteils auf die Teilkompetenzen **Testen von Modellen** und **Ändern von Modellen** bezogen: (1) Die Analysen in den Teilkompetenzen **Testen von Modellen** und **Ändern von Modellen** zeigten, dass für einen Großteil der Lernenden der Vergleich von Modell und Original zentral war. Vorherrschende Kategorien für die Teilkompetenz **Testen von Modellen** waren die Kategorie **Vergleich von Original und Modell** (33 % von $n_{\text{Lernende}} = 711$) und die Kategorie **Vergleich und**

Passung von Original und Modell (34 % von $n_{\text{Lernende}} = 711$) (Grünkorn, Upmeyer zu Belzen et al., 2014). Für die Teilkompetenz **Ändern von Modellen** war die Kategorie **Ändern bei mangelnder Passung mit dem Original** mit 61 % (von $n_{\text{Lernende}} = 712$) zentral. Auffällig war hierbei, dass Lernende beim Parallelisieren von Original und Modell den Zweck des Modells nicht beachteten. Der Zweck der Modelle wurde allerdings in den Testaufgaben beschrieben. So antworteten Lernende beispielsweise in der Teilkompetenz **Testen von Modellen** auf die Frage, wie man das Libellen-Modell testen könne, das für die Untersuchung der Flugfähigkeit von Libellen gebaut wurde:

„Man müsste überprüfen, ob das Modell realistisch und detailgetreu im Aussehen ist“ (Schüler 469).

Ähnliche Antworten zeigten sich auch in der Teilkompetenz **Ändern von Modellen**:

„Das Modell der Libelle kann noch verändert werden, weil es nicht perfekt wie eine Libelle aussieht. Es ist nicht originalgetreu“ (Schüler 970).

(2) Innerhalb der Teilkompetenz **Testen von Modellen** beschrieben Lernende sowohl abstrakte (Kategorie **Überprüfung von Hypothesen**) als auch konkrete (Kategorie **Überprüfung von Hypothesen mit Forschungsdesign**) Ideen. Schüler 253 antwortete beispielsweise:

„[...] Das Modell der Libelle ist somit vorteilhaft, da die Forscher es als Hilfe benutzen können und direkt auch Ideen mit in das Modell einbringen können. Diese Ideen könnten sie dann untersuchen und etwas über die Libelle erfahren. [...]“.

Hier beschreibt der Schüler allgemeine Ideen für Untersuchungen, um Hypothesen über das Original mit dem Modell überprüfen zu können. Auf eine konkrete Hypothese oder konkrete Anwendung geht der Schüler nicht ein. Im Vergleich dazu schrieb Schüler 413:

„Mit dem Modell könnte man seine Hypothesen überprüfen also zum Beispiel, ob die Fühler des Käfers empfindlich sind. Dafür könnte man dem Käfer Gegenstände in den Weg stellen, um zu sehen, woran er sich orientiert. Wenn er mit seinen Fühlern fühlt, könnte man versuchen ihn zu stören, sodass er seinen Weg wechselt. Wenn er dies tut, hat man bewiesen, dass der echte Käfer ausgeprägte Fühler hat.“

Dieser Schüler beschreibt eine konkrete Anwendung des Modells, mithilfe dessen eine Hypothese über das Original geprüft werden kann. Ähnliche Kategorien konnten auch in der Teilkompetenz **Zweck von Modellen** identifiziert werden: **Modell zum Überprüfen von abstrakten Ideen** und **Modell zum Überprüfen von konkreten Ideen**.

(3) Weiterhin erwähnten Lernende zum Teil in ihren Antworten eine Schritt-für-Schritt-Prozedur, die sie beim Testen oder Ändern einsetzen würden. Dabei kombinierten sie Kategorien aus verschiedenen Niveaus. Schülerin 186 antwortete beispielsweise auf eine Testaufgabe der Teilkompetenz **Testen von Modellen**:

„Als erstes überprüft man, ob das Modell überhaupt fliegen kann. Dann schaut man, welche Oberflächenstruktur vorhanden ist und vergleicht sie mit dem Modell. Danach schaut man, ob alle Bestandteile vorhanden sind. Wenn das Ok ist, stellt man Vermutungen über die Orientierung des Käfers an [...].“

Diese Schülerin überprüft zunächst, ob die Voraussetzungen für den Einsatz des Käfer-Modells erfüllt sind (Niveau I, Kategorie **Überprüfung der Grundvoraussetzungen**). Ein bewertender Vergleich des Modells mit dem Original erfolgt im zweiten und dritten Satz (Niveau II, **Kategorie Vergleich und Passung von Original und Modell**). Die Aussage der Schülerin endet mit dem Hinweis, dass mit dem Modell Vermutungen über das Original geprüft werden (Niveau III, Kategorie **Überprüfung von Hypothesen**).

Eine vergleichbare Vorgehensweise konnte auch für die Teilkompetenz **Ändern von Modellen** identifiziert werden. Schüler 492 äußerte sich beispielsweise:

„Das Modell wird verändert, wenn es erst gar nicht fliegt oder wenn es nicht wie eine Libelle fliegen kann“.

Für den Lernenden ist somit eine Änderung am Modell notwendig, wenn die Voraussetzungen für den Einsatz des Modells nicht erfüllt sind (Niveau I, Kategorie **Ändern bei Nichterfüllung der Grundvoraussetzungen**) und wenn das Modell nicht mit dem Original übereinstimmt (Niveau II, Kategorie **Ändern bei mangelnder Passung mit dem Original**).

5.2.3 Analyseergebnisse im Rahmen der IRT

Eine Voraussetzung für den Einsatz des entwickelten Leistungstests in einem probabilistischen Ansatz ist die Prüfung, inwiefern der Test in einem solchen Ansatz (z. B. Rasch-Skalierung) auch die dafür notwendigen Gütekriterien (z. B. Aufgaben-Fit) erfüllt (vgl. Kauertz, 2014). Diese Prüfung bezieht für das ordinale, mehrdimensionale Rasch-Modell – das Partial-Credit-Modell (Masters, 1982) – ähnlich wie bei den von Neumann (2014) beschriebenen Schritten Ergebnisse der *Prüfung der Modellanpassung*, der *Aufgabenparameter* und der *Personenfähigkeiten* ein (vgl. S. 91ff. dieser Arbeit).

Prüfung der Modellanpassung

Die Aufgaben im offenen Antwortformat wurden in ein-, zwei- und fünfdimensionalen IRT-Skalierungen mit Acer ConQuest 3.0.1 (Wu et al., 2012) geprüft. Die ein- und zweidimensionalen IRT-Skalierungen waren erfolgreich. Das bedeutet, dass die Schätzungen konvergiert sind und somit eine optimale Lösung gefunden wurde. Daher ist anzunehmen, dass die Anpassung eines Raschmodells an die Daten erfolgreich war.

Dagegen war für eine fünfdimensionale IRT-Skalierung die Modellanpassung an die Daten nicht erfolgreich. Dies ließ sich daraus ableiten, dass ConQuest die Warnung produzierte, die Schätzungen konvergierten zwar, das Ergebnis beruhe aber nicht auf einer optimalen Schätzung. Aus diesem Grund können keine belastbaren Analysen folgen und demzufolge kann der Leistungstest nicht geprüft werden (vgl. Neumann, 2014). Weiterhin ist zu bedenken, dass mit steigender Komplexität bzw. Dimensionalität des Modells mehr als fünfzehn Aufgaben benötigt werden, um verlässlich schätzen und zuverlässige Aussagen treffen zu können (Rost, 2004).

Demzufolge werden hier nur Ergebnisse der eindimensionalen IRT-Skalierung präsentiert. Die Kennwerte zur zweidimensionalen IRT-Skalierung sind in Anhang 5 nachzulesen. Für die fünfdimensionale IRT-Skalierung konnten keine Kennwerte ermittelt werden.

Aufgabenparameter

Tabellen 11 und 12 geben für den Aufgabenpool die notwendigen Kennwerte für das eindimensionale Strukturmodell wieder. Die Kennwerte der Tabelle 11 beziehen sich auf die Aufgabenebene, die Kennwerte der Tabelle 12 auf die

Ebene der Antwortkategorien (0, 1, 2 und 3). Die markierten Felder in Tabelle 12 kennzeichnen Auffälligkeiten, die im anschließenden Kapitel diskutiert werden (vgl. S. 123ff. dieser Arbeit).

Tab. 11: Kennwerte der eindimensionalen Raschskalierung auf Aufgabenebene

Aufgabe*	Bearbeitungen	Trennschärfe ^a	Schwierigkeit ^b	wMNSQ	t-Werte
EM1: Neandertaler	230	.56	0.344	1.03	0.3
EM2: Dinosaurier	232	.46	0.235	1.05	0.5
EM3: Biomembran	235	.43	0.719	0.99	-0.1
A1: Biomembran	235	.85	0.417	1.02	0.2
A2: Speiseröhre	235	.74	0.764	1.01	0.2
A3: Zunge	235	.75	0.398	0.99	-0.1
Z1: Wald	235	.69	-1.148	0.98	-0.2
Z2: Meer	235	.79	-0.859	1.02	0.2
Z3: Darm	236	.63	-0.527	0.97	-0.4
T1: Libelle	237	.68	-0.175	1.04	0.4
T2: Zanonía	237	.65	-0.057	1.01	0.1
T3: Käfer	237	.66	0.058	0.98	-0.2
Ä1: Libelle	237	.67	1.154	0.99	-0.1
Ä2: Krabbe	238	.88	1.122	1.03	0.3
Ä3: Mund	237	.75	0.984	0.97	-0.3

Anmerkungen. * Die Buchstaben kennzeichnen die Teilkompetenz der Aufgabe (EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen). Die Zahl kennzeichnet die Nummer der Aufgabe, zum Beispiel Z1 = Aufgabe 1, der Teilkompetenz **Zweck von Modellen**

^a Für die Berechnung der Trennschärfen wurden die Aufgaben pro Teilkompetenz eindimensional skaliert.

^b Die Werte der Aufgabenschwierigkeit beruhen auf Mittelwerten der Thresholdparameter.

Tab. 12: Kennwerte der eindimensionalen Raschskalierung auf Ebene der Antwortkategorien (0, 1, 2 und 3)

Aufgabe*	Häufigkeiten (%)				wMNSQ				t-Werte				Punktbiseriale Korrelationen				Item-Deltas (δ)			Item Thresholds (τ) ^a		
	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3	δ_1	δ_2	δ_3	τ_1	τ_2	τ_3
EM1	9.13	71.30	14.78	4.78	1.02	1.01	1.00	1.05	0.2	0.1	0.0	0.2	-0.12	-0.05	0.10	0.11	-2.34	1.71	1.67	-2.36	1.24	2.16
EM2	6.90	65.52	22.84	4.74	1.00	1.03	1.01	1.05	0.1	0.5	0.1	0.3	-0.09	-0.08	0.15	0.00	-2.53	1.17	2.07	-2.56	0.92	2.34
EM3	14.89	69.36	12.77	2.98	1.00	1.00	0.98	1.00	0.0	-0.1	-0.1	0.1	-0.17	0.01	0.15	0.05	-1.76	1.88	2.03	-1.78	1.47	2.48
A1	35.74	28.09	22.13	14.04	0.98	0.99	0.99	1.07	-0.3	-0.1	-0.1	0.6	-0.28	-0.01	0.22	0.14	0.01	0.35	0.89	-0.44	0.39	1.29
A2	24.26	36.60	35.32	3.83	0.98	0.99	0.98	1.04	-0.2	-0.1	-0.4	0.2	-0.20	-0.07	0.27	-0.04	-0.63	0.18	2.74	-0.92	0.40	2.81
A3	19.15	41.70	31.91	7.23	1.00	0.99	0.98	1.01	0.0	-0.2	-0.4	0.1	-0.10	-0.18	0.23	0.08	-1.08	0.34	1.93	-1.27	0.37	2.10
Z1	1.28	25.11	49.79	23.83	1.01	0.99	0.99	1.00	0.2	-0.1	-0.7	0.0	-0.13	-0.20	0.12	0.10	-3.54	-0.85	0.94	-3.60	-0.92	1.08
Z2	3.40	27.66	41.70	27.23	1.04	0.97	0.99	1.04	0.2	-0.5	-0.2	0.5	-0.07	-0.23	0.20	0.03	-2.64	-0.57	0.64	-2.75	-0.68	0.86
Z3	5.93	26.27	47.46	20.34	0.99	0.96	0.98	1.00	0.0	-0.5	-0.8	0.0	-0.14	-0.26	0.13	0.20	-1.98	-0.71	1.11	-2.19	-0.64	1.24
T1	7.59	19.41	62.87	10.13	1.03	1.03	1.01	0.98	0.2	0.3	0.3	-0.1	-0.20	-0.04	0.06	0.13	-1.42	-1.26	2.16	-1.86	-0.86	2.19
T2	7.17	24.89	61.18	6.75	1.04	0.98	0.99	0.99	0.3	-0.2	-0.3	0.0	-0.09	-0.22	0.18	0.12	-1.71	-0.98	2.52	-2.02	-0.71	2.55
T3	10.13	16.88	66.24	6.75	0.95	1.00	0.98	1.03	-0.3	0.1	-0.4	0.2	-0.23	-0.10	0.25	-0.05	-0.97	-1.44	2.59	-1.59	-0.84	2.60
Ä1	14.77	18.57	66.24	0.42	0.96	0.99	1.00	1.04	-0.3	0.0	0.1	0.4	-0.20	-0.01	0.17	-0.04	-0.67	-1.31	5.45	-1.35	-0.64	5.45
Ä2	13.87	17.23	68.49	0.42	1.06	0.98	1.00	1.00	0.5	-0.2	-0.1	0.3	-0.08	-0.15	0.17	0.10	-0.67	-1.44	5.47	-1.39	-0.72	5.47
Ä3	15.19	14.35	69.62	0.84	1.00	0.98	0.95	0.99	0.0	-0.1	-0.7	0.2	-0.18	-0.17	0.27	-0.01	-0.32	-1.57	4.84	-1.21	-0.68	4.84

Anmerkungen. Unter Antwortkategorie 0 sind unbeantwortete Aufgaben (Rekodierung fehlender Daten, vgl. Rost, 2004, S. 87), nicht mit der Frage korrespondierende Antworten sowie Antworten auf dem basalen Niveau subsumiert. Die Antwortkategorien 1, 2 und 3 entsprechen den Niveaus des Kompetenzmodells. **Orange markierte Felder** kennzeichnen entweder Antwortkategorien mit Häufigkeiten, die unter zehn Prozent liegen, oder negative punktbiseriale Korrelationen für die höchste Antwortkategorie oder ungeordnete Schwellenparameter (Item-Deltas). **Grün markierte Felder** kennzeichnen Aufgaben mit einer noch akzeptablen Veränderung von negativen für niedrige Antwortkategorien zu positiven punktbiserialen Korrelationen für höhere Antwortkategorien.

* Die Buchstaben kennzeichnen die Teilkompetenz der Aufgabe (EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen). Die Zahl kennzeichnet die Nummer der Aufgabe, zum Beispiel Z1 = Aufgabe 1, der Teilkompetenz **Zweck von Modellen**.

^a Diejenige Stelle auf der Skala, an der die Wahrscheinlichkeit die Antwortkategorie k oder einer höhere zu erzielen 50 Prozent beträgt.

Zur Verdeutlichung der geordneten und ungeordneten Schwellen wird in Abbildung 10 eine Beispielaufgabe mit geordneten Schwellen dargestellt.

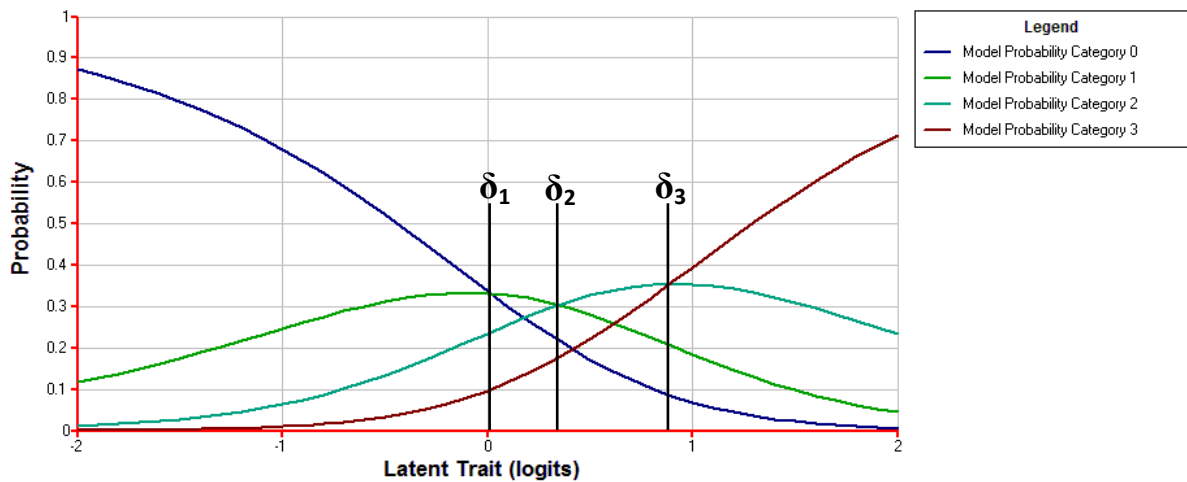


Abb. 10: Characteristic Curves by Score mit geordneten Schwellen ($\delta_1 = 0.01$; $\delta_2 = 0.35$; $\delta_3 = 0.89$) am Beispiel der Aufgabe A1

Im Gegensatz dazu zeigt Abbildung 11 eine Beispielaufgabe mit ungeordneten Schwellen.¹⁰ Bei diesem Beispiel liegt der Schnittpunkt zwischen den Kategorienfunktionen 1 und 2 ($\delta_2 = -1.57$) vor dem Schnittpunkt der Kategorienfunktionen 0 und 1 ($\delta_1 = -0.32$).

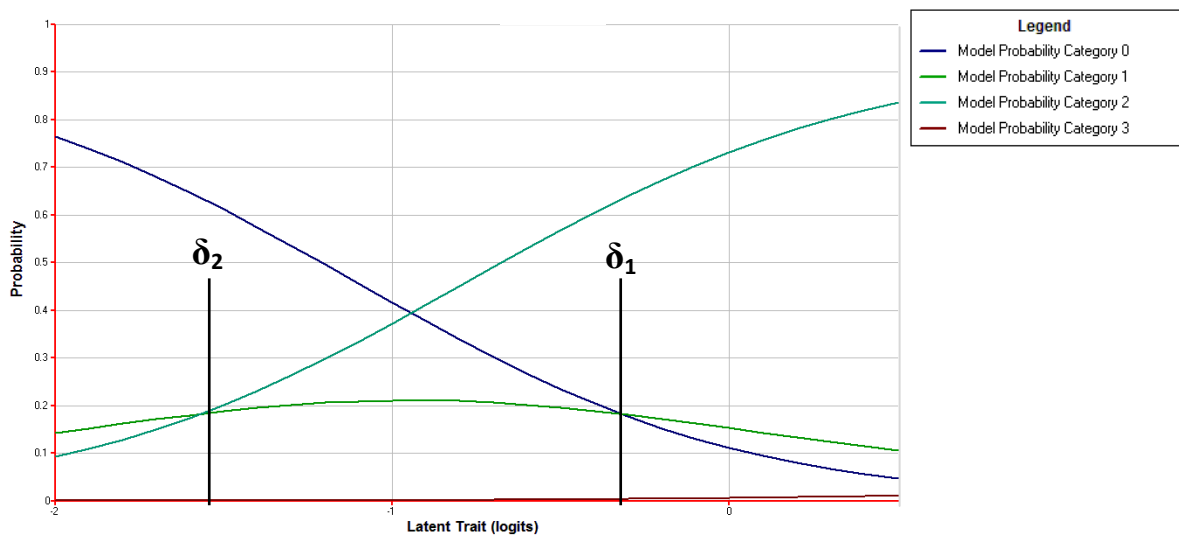


Abb. 11: Ausschnitt der Characteristic Curves by Score mit ungeordneten Schwellen ($\delta_1 = -0.32$; $\delta_2 = -1.57$)¹⁰ am Beispiel der Aufgabe Ä3

¹⁰ Aus Gründen der Übersichtlichkeit zeigt die Abbildung 11 nur einen Ausschnitt. Der dritte Schnittpunkt zwischen den Antwortkategorien 2 und 3 ist nicht dargestellt ($\delta_3 = 4.84$).

Personenfähigkeiten

Mit Blick auf die Verteilung der Personenparameter gibt Tabelle 13 einen Überblick über die Kennwerte. Auch hier sind entsprechenden Auffälligkeiten gekennzeichnet.

Tab. 13: Mittelwerte der Personenfähigkeiten auf Ebene der Antwortkategorien (0, 1, 2 und 3)

Aufgabe*	Mittelwerte bei den Personenfähigkeiten auf Ebene der Antwortkategorien (0, 1, 2 und 3)			
	0	1	2	3
EM1	-0.55	-0.06	0.42	0.42
EM2	-0.42	-0.04	0.36	0.09
EM3	-0.45	0.09	0.44	0.65
A1	-0.44	0.02	0.35	0.52
A2	-0.40	-0.10	0.39	0.42
A3	-0.51	-0.10	0.33	0.55
Z1	-0.72	-0.31	0.10	0.37
Z2	-0.65	-0.31	0.11	0.32
Z3	-0.70	-0.33	0.13	0.47
T1	-0.80	-0.33	0.10	0.50
T2	-0.67	-0.37	0.15	0.58
T3	-0.61	-0.26	0.18	0.48
Ä1	-0.73	-0.19	0.15	1.09
Ä2	-0.44	-0.25	0.14	1.00
Ä3	-0.51	-0.28	0.24	0.73

Anmerkungen. Unter Antwortkategorie 0 sind unbeantwortete Aufgaben (Rekodierung fehlender Daten, vgl. Rost, 2004, S. 87), nicht mit der Frage korrespondierende Antworten sowie Antworten auf dem basalen Niveau subsumiert. Die Antwortkategorien 1, 2 und 3 entsprechen den Niveaus des Kompetenzmodells. Orange markierte Felder kennzeichnen Mittelwerte der Personenfähigkeiten, die nicht aufsteigen.

* Die Buchstaben kennzeichnen die Teilkompetenz der Aufgabe (EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen). Die Zahl kennzeichnet die Nummer der Aufgabe, zum Beispiel Z1 = Aufgabe 1, der Teilkompetenz **Zweck von Modellen**.

Neben den Aufgabenparametern und den Mittelwerten der Personenfähigkeiten sind zur Beurteilung des Tests die Varianz der Personenverteilung und die EAP/PV-Reliabilität zu ermitteln sowie die Wright Map heranzuziehen. Die Varianz beträgt 0.476, die EAP/PV-Reliabilität .398. Abbildung 12 präsentiert die Wright Map der eindimensionalen Skalierung und skizziert die Verteilung der Personenparameter (links) sowie die Verteilung des Aufgabenparameters **Thurstonian Thresholds** (rechts) getrennt nach Teilkompetenzen.

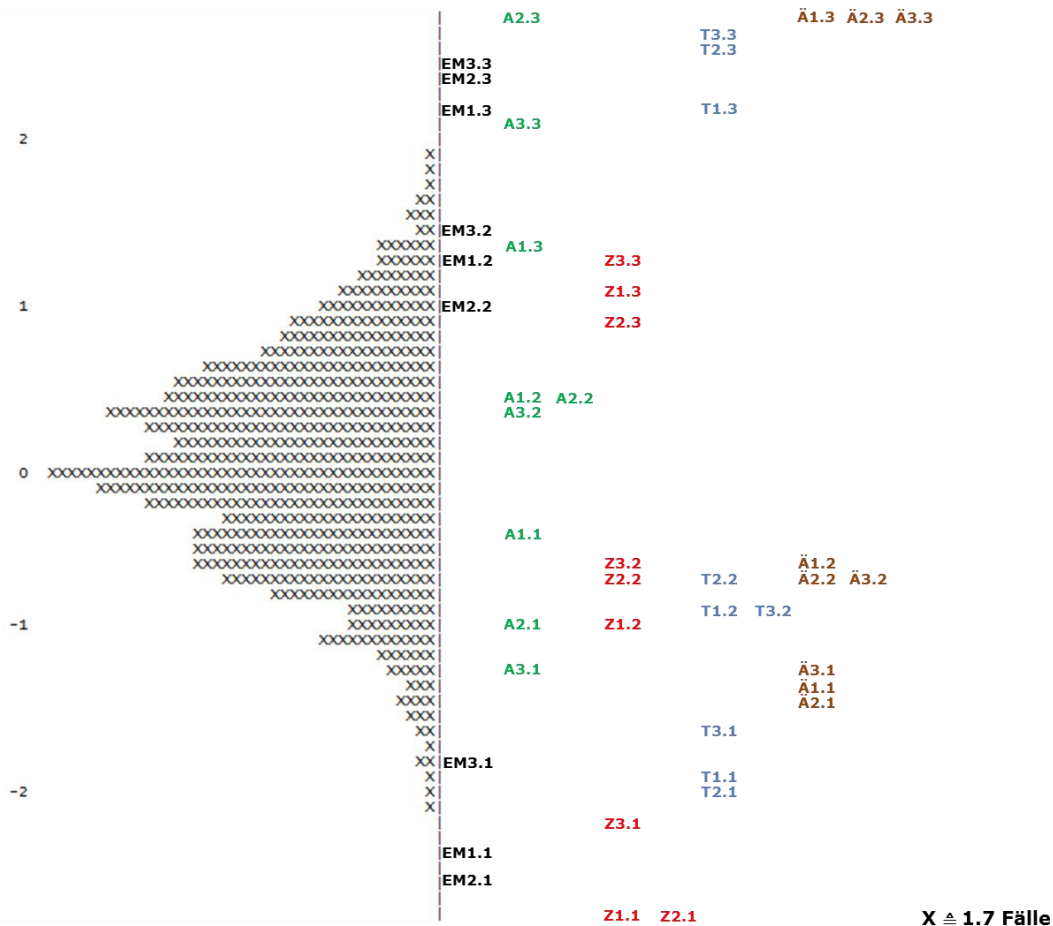


Abb. 12: Wright Map der eindimensionalen Skalierung mit den Personenfähigkeiten (links) und den Thurstonian Thresholds (rechts)¹¹

5.3 Zusammenfassende Diskussion

Auf der Grundlage der Analysen wird zusammenfassend diskutiert, ob das Kompetenzmodell der Modellkompetenz die Perspektiven der Lernenden auf Modelle hinreichend reflektiert (Beitrag 3), welche Besonderheiten in den Antwortstrukturen der Lernenden hilfreiche Erklärungshinweise für die Perspektiven der Lernenden auf Modelle geben und inwiefern der Leistungstest auch in einem probabilistischen Ansatz geeignet ist und eingesetzt werden kann.

¹¹ Die Buchstaben kennzeichnen die Teilkompetenz der Aufgabe (EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen). Die erste Zahl kennzeichnet die Nummer der Aufgabe und die zweite Zahl den Thurstonian Threshold, z. B. Z1.1 = Aufgabe 1, Threshold 1 (Schwelle an der es wahrscheinlicher ist, Kategorie 1 anstatt Kategorie 0 zu wählen) der Teilkompetenz **Zweck von Modellen**.

5.3.1 Inhaltliche Überprüfung der theoretischen Struktur (Beitrag 3)

Die Auswertung der Antworten der Lernenden ergab, dass die Perspektiven von Lernenden der Sekundarstufe I auf die Teilkompetenzen *Eigenschaften von Modellen* und *Zweck von Modellen* umfassend und adäquat beschrieben werden (vgl. Upmeier zu Belzen & Krüger, 2010). Die vorliegende Arbeit repliziert damit Befunde von Crawford und Cullin (2004), Grosslight et al. (1991) sowie Justi und Gilbert (2003). In den Teilkompetenzen *Alternative Modelle*, *Testen von Modellen* und *Ändern von Modellen* wurden im Vergleich zu Upmeier zu Belzen und Krüger (2010) und teilweise auch im Vergleich zu Crawford und Cullin (2004), Grosslight et al. (1991) sowie Justi und Gilbert (2003) zusätzliche Perspektiven von Lernenden identifiziert. Damit stimmen die Ergebnisse dieser Studie mit Befunden aus den vorangegangenen Teilstudien 1 und 2 überein (vgl. Grünkorn & Krüger, 2012; Grünkorn et al., 2011).

Die Teilkompetenz *Alternative Modelle* basiert auf dem Verständnis, dass es zu einem Phänomen mehrere Modelle geben kann (vgl. Upmeier zu Belzen & Krüger, 2010). Einige Lernende begründeten das Vorhandensein der drei gezeigten Modelle zu einem Phänomen in den Testaufgaben damit, „dass es völlig sinnlos [ist], weil alle dasselbe darstellen“ (Schüler 415) (*Modelle sind gleich*), dass es verschiedene Modelle gebe, weil nicht jede Zunge gleich sei (Schülerin 1185) (*Verschiedene Modelle zu unterschiedlichen Originalen*) und, dass „die anderen [Modelle] falsch sind und es nur ein richtiges Modell gibt“ (Schüler 643) (*Nur ein endgültiges und richtiges Modell*). Diesen Begründungen liegt das Verständnis zugrunde, dass es zu einem Phänomen nur ein Modell geben kann (übergeordnete Kategorie: *Nur ein Modell zu einem Modell*). Vor diesem Hintergrund können die identifizierten Kategorien nicht den Niveaus der Teilkompetenz zugeordnet werden.

Auch andere Studien liefern vereinzelt Hinweise auf die Kategorien *Modelle sind gleich* und *Nur ein endgültiges und richtiges Modell*. Grosslight et al. (1991) weisen in ihrer Studie darauf hin, dass einige Lernende die Vorstellung zeigten, es sei nicht möglich, zu einem Phänomen verschiedene Modelle zu haben. Justi und Gilbert (2003) beschrieben in ihrer Studie mit Lehrkräften diesen Aspekt explizit (*uniqueness*, S. 1375), der mit den identifizierten Kategorien *Modelle sind gleich* und *Nur ein endgültiges und richtiges Modell* vergleichbar ist. Bei Lernenden, welche die letztere Begründung heranzogen, ist

anzunehmen, dass sie diese Modelle als endgültige Versionen wahrnahmen und deshalb die verschiedenen Modelle auch auf ihre Korrektheit hin beurteilten. Im Gegensatz dazu fragen Forschende aus einer wissenschaftlichen Perspektive heraus nicht danach, ob Modelle richtig oder falsch sind, sondern nur, ob das wahrgenommene Phänomen mit dem Modell erklärt werden kann (Crawford & Cullin, 2004). Auch die Kategorie **Verschiedene Modelle zu unterschiedlichen Originalen** stellt eine Ergänzung zu den Befunden von Crawford und Cullin (2004), Grosslight et al. (1991) sowie Justi und Gilbert (2003) dar. Vor diesem Hintergrund ist festzuhalten, dass die übergeordnete Perspektive **Nur ein Modell zu einem Original** mit ihren drei Kategorien unter Schülerinnen und Schüler vorherrscht. Daher sollte sie als basales Niveau im Kompetenzmodell der Modellkompetenz berücksichtigt werden.

Auch für die Teilkompetenz **Testen von Modellen** konnte eine zusätzliche Kategorie (**Keine Testung des Modells**) identifiziert werden. Dies äußerte sich darin, dass Lernende ein Testen von Modellen als nicht notwendig erachteten. Dessen Akzeptanz ist jedoch grundlegend für die Teilkompetenz **Testen von Modellen** (vgl. Upmeier zu Belzen & Krüger, 2010). Während die Studien von Crawford und Cullin (2004), Grosslight et al. (1991) sowie Justi und Gilbert (2003) keine Hinweise auf diese Kategorie lieferten, kamen Hänsch und Upmeier zu Belzen (2012; Grünkorn, Hänsch, Upmeier zu Belzen & Krüger, 2012) in einer Studie mit Hands-on-Aufgaben zu ähnlichen Ergebnissen. Aus psychometrischer Sicht kann diese Kategorie aufgrund des geringen Vorkommens (1 %) vernachlässigt werden. Aus fachdidaktischer Sicht ist sie insofern bedeutend, als sie von Lehrkräften für eine adäquate Förderung der Lernenden berücksichtigt werden müsste. Erst wenn Lehrkräfte wissen, dass Lernende Testungen von Modellen ablehnen, können sie sich im Unterricht darauf einstellen und gegebenenfalls angemessene Maßnahmen erarbeiten (vgl. S. 137ff. dieser Arbeit). Daher wird die Kategorie **Keine Testung des Modells** als basales Niveau vorgeschlagen.

Auch in Bezug auf die Teilkompetenz **Ändern von Modellen** zeigte sich in der Analyse analog zum **Testen von Modellen**, dass einige Lernende ein Ändern von Modellen grundsätzlich ablehnen. Diese Perspektive auf Modelle wurde bereits in anderen Studien festgestellt (Crawford & Cullin, 2004, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Treagust et al., 2002). Da die Akzeptanz des Ändern von Modellen für diese Teilkompetenz grundlegend ist, kann diese Kategorie keinem Niveau der Teilkompetenz zugeordnet werden (vgl.

Upmeier zu Belzen & Krüger, 2010). Über die Kategorie *Kein Anlass für eine Änderung* hinaus begründen manche Lernende Änderungen am Modell mit unterschiedlichen Originalen, die dargestellt werden sollen (Kategorie *Ändern zur Darstellung unterschiedlicher Originale*). Ähnlich wie bei der Kategorie *Verschiedene Modelle zu unterschiedlichen Originalen* der Teilkompetenz *Alternative Modelle* beziehen sich die Lernenden auch bei dieser Kategorie auf unterschiedliche Originale. Die Entwicklung, Überprüfung und Änderung eines Modells fokussiert allerdings immer auf ein Original und ist zweckgebunden (vgl. Giere, 2001, 2010; Halloun, 2001; Mahr, 2008; Stachowiak, 1973). Beide Kategorien (*Kein Anlass für eine Änderung* und *Ändern zur Darstellung unterschiedlicher Originale*) beschreiben ein basales Verständnis, das Änderungen am Modell ablehnt. Daher sollten diese Kategorien im Kompetenzmodell als basales Niveau berücksichtigt werden.

Neben diesen basalen Kategorien beschreiben Schülerinnen und Schüler auch weitere Kategorien in den Teilkompetenzen *Zweck* und *Ändern von Modellen* (vgl. Tab. 8). Diese Kategorien beziehen sich auf alltägliche Perspektiven auf Modelle, die für den erkenntniswissenschaftlichen Prozess keine Rolle spielen (vgl. Coll et al., 2005; Odenbaugh, 2005). Es ist anzunehmen, dass sie für die Entwicklung der Modellkompetenz wenig bedeutsam sind, deshalb erscheint es nicht sinnvoll, sie im Kompetenzmodell zu berücksichtigen.

Auf der Basis dieser Ergebnisse kann die Hypothese, dass die theoretische Struktur des Kompetenzmodells der Modellkompetenz die Perspektiven von Lernenden auf Modelle und auf die Modellbildung hinreichend beschreibt, teilweise bestätigt werden (H1). Crawford und Cullin (2005) definierten beispielsweise für die Aspekte *purpose of models*, *changing a model* und *multiple models for the same thing* vier Niveaus. Damit differenzierten sie diese Aspekte stärker als Upmeier zu Belzen und Krüger (2010) es für ihre Teilkompetenzen *Alternative Modelle*, *Zweck von Modellen* und *Ändern von Modellen* tun. Die vorliegenden Befunde lassen die differenzierte Beschreibung von Crawford und Cullin (2005) in vier Niveaus für die Teilkompetenzen *Alternative Modelle*, *Testen von Modellen* und *Ändern von Modellen* plausibel erscheinen.

Im Hinblick auf die Niveaus, welche die Lernenden dieser Stichprobe erreichten, zeigte sich, dass die meisten Lernenden Niveau I und II, aber nur wenige Niveau III erreichten. Dieses Ergebnis ist erwartungskonform (H2) und mit dem anderer nationaler und internationaler Studien vergleichbar (e. g., Gross-

light et al., 1991; Schwarz & White, 2005; Trier & Upmeier zu Belzen, 2009). Unterschiede bei den Häufigkeiten der Kategorien konnten vor allem in den Teilkompetenzen *Alternative Modelle* und *Ändern von Modellen* festgestellt werden. Während bei Grosslight et al. (1991) häufig unterschiedliche inhaltliche Schwerpunkte als Begründung für alternative Modelle genannt wurden, wurden in der vorliegenden Studie vor allem unterschiedliche Modellobjektigenschaften als Begründung herangezogen. Darüber hinaus war in dieser Studie die Kategorie *Ändern bei neuen Erkenntnissen über das Original* in der Teilkompetenz *Ändern von Modellen* weniger zentral als bei Grosslight et al. (1991).

Ein wesentlicher Grund für diese unterschiedlichen Häufigkeiten könnte darin liegen, dass in der vorliegenden Arbeit ein kontextabhängiger Ansatz gewählt wurde und die Aufgaben entsprechend in verschiedene biologische Kontexte eingebettet waren. Sins et al. (2009) wiesen in ihrer Studie darauf hin, dass kontextunabhängige und kontextabhängige Ansätze unterschiedliche Ergebnisse hervorbringen können. Krell, Upmeier zu Belzen et al. (2012b, 2014c) stellten in ihrer Studie zu FC-Items fest, dass sich die Ergebnisse innerhalb eines kontextabhängigen Ansatzes unterscheiden können, und Terzer (2013; Terzer, Patzke & Upmeier zu Belzen, 2012) unterstützt dies in ihrer Studie zu MC-Items.

Insgesamt ist festzuhalten, dass die Befragten in dieser Stichprobe Modelle vorwiegend aus einer medialen Sicht betrachteten und somit weniger als ein Mittel der Erkenntnisgewinnung wahrnahmen (vgl. Mahr, 2009; Oh & Oh, 2011). Geht man wie Crawford und Cullin (2005) davon aus, dass das Modellverständnis der Lehrkräfte das der Lernenden beeinflusst, könnten diese vorherrschenden Perspektiven auf Modelle darauf hinweisen, dass Modelle im Unterricht selten als Mittel der Voraussage und Generierung von Hypothesen eingesetzt werden und dass der Umgang mit und das Denken in Modellen kaum thematisiert werden (van Driel & Verloop, 2002).

5.3.2 Besonderheiten in den Antwortstrukturen der Lernenden

Die Analyse der Antworten der Lernenden führte zu den Kategorien *Modell als Mittel der Zugänglichkeit* und *Modell als Mittel der Verständlichkeit/Kommunizierbarkeit*. Damit unterstützt dieser Befund die Ergebnisse aus der Studie von Grosslight et al. (1991) mit Schülerinnen und Schülern und aus

der Studie von Ingham und Gilbert (1991) mit Studierenden der Chemie (H3). Auch Orsenne und Upmeier zu Belzen (2012) beschrieben in ihrer Studie zur Erfassung und Förderung von Modellkompetenz mit Hands-on Aufgaben die **Kategorie „Erkennen eines sozialen Aspekts“ (S. 40), die mit der Kategorie *Modell als Mittel der Verständlichkeit/Kommunizierbarkeit* vergleichbar ist.**

Ergänzend zu diesen Studien war eine differenzierte Kodierung in die beiden didaktisch orientierten Perspektiven möglich, die es erlaubt, zu überprüfen, in welchen Teilkompetenzen sie auftreten und ob sie mit einer medialen oder methodischen Perspektive (bzw. Niveaus) in Verbindung stehen. Die Ergebnisse zeigen, dass die Kategorien ***Modell als Mittel der Zugänglichkeit*** und ***Modell als Mittel der Verständlichkeit/Kommunizierbarkeit*** bei Antworten in allen Teilkompetenzen des Kompetenzmodells identifiziert werden konnten. Darüber hinaus konnten die Kategorien in Verbindung mit einer medialen und methodischen Sichtweise auf Modelle beschrieben werden (H4). Dadurch, dass diese Perspektiven sowohl in den fünf Teilkompetenzen als auch in Verbindung mit einer medialen und methodischen Sichtweise auf Modelle auftraten, ist anzunehmen, dass zwischen den beiden Perspektiven und den Teilkompetenzen der Modellkompetenz Zusammenhänge bestehen. Die Untersuchung dieser Zusammenhänge könnte Aufschluss darüber geben, ob die beiden didaktisch orientierten Perspektiven eine Schlüsselrolle bei der Förderung und Entwicklung von Modellkompetenz einnehmen können (vgl. S. 145ff. dieser Arbeit).

Dass insbesondere der Kommunikations- und Verständnisaspekt eine wichtige Rolle bei der Entwicklung eines elaborierten Modellverständnisses spielt, verdeutlichen Schwarz et al. (2009) dadurch, dass sie in ihrem Strukturierungsansatz diesem Aspekt eine eigene Dimension (***sensemaking and communication***) zuweisen. Sie argumentieren, dass den Lernenden und den Forschenden mithilfe der Kommunikation nicht nur Wissen vermittelt und damit ein besseres Verständnis über Phänomene ermöglicht wird, sondern ihnen auch der Austausch und die Diskussion von Ideen nähergebracht werden kann. Durch eine intensive Reflexion der beide didaktisch orientierten Kategorien können im **Unterricht die „Anschauungsfunktion“, die „denkökonomische Funktion“ und die „heuristische Funktion“ (vgl. Kattmann, 2006, S. 332) von Modellen in den Blick genommen werden. Zudem kann den Lernenden das Potenzial von Modellen als Mediatoren, die zwischen Theorie und Phänomen vermitteln (vgl. Oh & Oh, 2011), verdeutlicht werden.**

Im Hinblick auf die Besonderheiten in den Teilkompetenzen **Testen von Modellen** und **Ändern von Modellen** zeigten die Ergebnisse, dass der Vergleich von Modell und Original zentral für Schülerinnen und Schüler ist. Dieses Ergebnis war nach den Befunden aus den Teilstudien 1 und 2 (Grünkorn & Krüger, 2012) zu erwarten, ist allerdings, wie weiter oben bereits diskutiert, mit den Ergebnissen der Studie von Grosslight et al. (1991) nur teilweise vergleichbar. Eine ähnliche Besonderheit beschreibt Terzer (2013; Terzer et al., 2012) in einer Studie zur Untersuchung der kognitiven Prozesse von Lernenden bei der Bearbeitung von MC-Items mit der Methode des lauten Denkens. In den Denkprotokollen der Lernenden zeigte sich, dass der Vergleich von Modell und Original ebenfalls zentral war und häufig zur Beantwortung der Aufgaben herangezogen wurde. Ausgehend von diesen Befunden ist anzunehmen, dass das Parallelisieren von Modell und Original bei den Schülerinnen und Schülern sehr präsent und basal verankert ist (vgl. S. 145ff. dieser Arbeit).

Eine zusätzliche Erklärung ist, dass das Parallelisieren vermutlich intensiv im Unterricht angewandt wird. Für Letzteres spricht der Befund, dass die Schülerinnen und Schülern den Zweck des Modells beim Vergleich von Original und Modell unbeachtet lassen. Gerade die Zweckgebundenheit spielt jedoch bei der Konstruktion von Modellen eine große Rolle (Mahr, 2008; Stachowiak, 1973). Darüber, inwieweit das Parallelisieren im Unterricht stattfindet, könnte die Untersuchung von Lerngelegenheiten Aufschluss geben (vgl. Patzke & Upmeyer zu Belzen, 2012).

Neben dieser vorherrschenden Perspektive konnten in den Antworten der Lernenden zur Teilkompetenz **Testen von Modellen** konkrete und abstrakte Ideen für ein Testen unterschieden werden. Diese Unterscheidung fand sich auch in der Teilkompetenz **Zweck von Modellen**. Auch wenn Crawford und Cullin (2005), Grosslight et al. (1991) sowie Justi und Gilbert (2003) in ihren Studien keine vergleichbaren Kategorien beschrieben, wiesen Leisner-Bodenthin (2006) sowie Leisner und Mikelskis (2004) auf eine domänenspezifische und domänenübergreifende Modellkompetenz bzw. auf ein domänenspezifisches und domänenübergreifendes Modellwissen hin. Dieser Befund wird zudem durch Studien von Krell, Upmeyer zu Belzen et al. (2012b, 2014c) und von Terzer (2013) unterstützt. Krell, Upmeyer zu Belzen et al. (2012b) stellten Unterschiede in der Teilkompetenz **Zweck von Modellen** bei der Performanz von Lernenden fest, die kontextabhängigen und kontextunabhängigen FC-Items bear-

beiteten. Terzer (2013) identifizierte in Denkprotokollen von Schülerinnen und Schülern sowohl abstrakte als auch konkrete Vorstellungen von Modellen. Dabei zeigten die Schülerinnen und Schüler, dass sie zwar abstrakte Vorstellungen von Modellen formulieren können, diese aber nicht auf konkrete Situationen übertragen. Diese Befunde sprechen dafür, zwischen einem konkreten und einem abstrakten Modellverständnis zu unterscheiden und diese Unterscheidung bei der Förderung von Modellkompetenz zu berücksichtigen (vgl. Fleige et al., 2012b; Grünkorn & Fleige, 2012; Grünkorn, Lotz et al., 2014; vgl. S. 137ff. dieser Arbeit).

In Bezug auf den Prozess des Testens und Änderns von Modellen konstatierten Grosslight et al. (1991) und Treagust et al. (2002), dass Schülerinnen und Schüler diesen als unsystematisch ansehen. In den Antworten der Lernenden dieser Stichprobe konnte allerdings vereinzelt eine Schritt-für-Schritt-Prozedur identifiziert werden, die verschiedene Niveaus der Teilkompetenzen systematisch einbindet. Aufschluss über diese unterschiedlichen Befunde könnte ebenfalls die Untersuchung der Lerngelegenheiten geben (Patzke & Upmeyer zu Belzen, 2012).

5.3.3 Analysen im Rahmen der IRT

Für den Einsatz des Leistungstests in einem probabilistischen Ansatz lässt sich festhalten, dass die Kennwerte – Trennschärfe, Schwierigkeit, $wMNSQ$ und t -Wert – für die 15 Aufgaben (Tab. 11 und 12) in einem akzeptablen Bereich liegen (vgl. Bond & Fox, 2007; Bühner, 2010; Kelava & Moosbrugger, 2012; Neumann, 2014; Rost, 2004; Wu et al., 2007). Es ist auf der Basis dieser Kennwerte zunächst nicht notwendig, Aufgaben auszuschließen. Bei der Betrachtung der punktbiserialen Korrelationen, der Schwellenparameter (Item-Deltas) und der Verteilung der Personenparameter konnten jedoch auffällige Aufgaben identifiziert werden (vgl. Tab. 12 und 13). Diese auffälligen Aufgaben sind in Tabelle 14 zusammengefasst.

Tab. 14: Aufgaben mit negativen punktbiserialen Korrelationen für die höchste Kategorie, mit ungeordneten Schwellenparametern und mit abnehmenden Mittelwerten der Personenfähigkeiten

Aufgabe*	Punktbiserialer Korrelationen: negative Werte für die höchste Kategorie	Ungeordnete Schwellenparameter (δ_i)	Abnehmende Mittelwerte der Personenfähigkeiten
EM1: Neandertaler		$\delta_1 < \delta_2 > \delta_3$	X
EM2: Tyrannosaurus rex			X
A2: Speiseröhre	X		
T3: Käfer	X	$\delta_1 > \delta_2 < \delta_3$	
Ä1: Libelle	X	$\delta_1 > \delta_2 < \delta_3$	
Ä2: Mund		$\delta_1 > \delta_2 < \delta_3$	
Ä3: Krabbe	X	$\delta_1 > \delta_2 < \delta_3$	

Anmerkung. * Der Buchstabe kennzeichnet die Teilkompetenz der Aufgabe (EM = Eigenschaften von Modellen, A = Alternative Modelle, T = Testen von Modellen, Ä = Ändern von Modellen). Die Zahl kennzeichnet die Nummer der Aufgabe, zum Beispiel A1 = Aufgabe 1, der Teilkompetenz **Alternative Modelle**. **X** bedeutet, dass die Aufgabe diese Auffälligkeit zeigt.

In Bezug auf die *punktbiserialen Korrelationen* sind besonders negative Korrelationen für die höchste Antwortkategorie 3 problematisch. Aufgaben, für die nur ein ungleichmäßiger Anstieg zu verzeichnen ist und die für die niedrigste Antwortkategorie keine positiven und für die höchste Antwortkategorie keine negativen punktbiserialen Korrelationen zeigen, sind dagegen noch akzeptabel (Bühner, 2010; Hartig, 2009). Bei den Aufgaben A2, T3, Ä1 und Ä3 ist dies jedoch nicht der Fall. Bei diesen Aufgaben korreliert das Erreichen von Niveau III nicht gut mit der Gesamtleistung im Test bzw. repräsentiert diese nicht gut. So erreichen beispielsweise auch Schülerinnen und Schüler mit schlechter Gesamtleistung im Test das Niveau III dieser Aufgabe. Daraus lässt sich schließen, dass diese Aufgaben nicht gut zwischen Personen mit hoher Gesamtleistung und solchen mit niedriger Gesamtleistung trennen (vgl. Bühner, 2010; Rost, 2004).

Neben den punktbiserialen Korrelationen zeigten die Aufgaben EM1, T3, Ä1, Ä2 und Ä3 *ungeordnete Schwellenparameter* (vgl. Tab. 14). Solche Aufgaben sind dadurch charakterisiert, dass es für eine Kategorie keinen Abschnitt auf dem latenten Kontinuum gibt, an dem diese mit höchster Wahrscheinlichkeit von einer Person erreicht werden kann (Rost, 2004, S. 205). Somit liegen die Antwortkategorien nicht, wie angenommen, geordnet vor (Bühner, 2010; Rost, 2004).

Im Hinblick auf die Mittelwerte der Personenfähigkeiten erwartet man bei ordinalen Skalen *Mittelwerte, die mit den Antwortkategorien ansteigen*. Dieser An-

stieg konnte bei den Aufgaben EM1 und EM2 nicht festgestellt werden. Hier stiegen die Mittelwerte der Personenfähigkeiten nur bis zur Antwortkategorie 2 an: Beim Übergang von Antwortkategorie 2 auf 3 blieben die Mittelwerte der Personenfähigkeiten entweder gleich (EM1) oder sie nahmen ab (EM2). Somit sind mit dem Erreichen höherer Antwortkategorien nicht auch höhere Personenfähigkeiten verbunden. Damit haben beispielsweise auch Personen im Mittel mit geringeren Fähigkeiten das Niveau III der Aufgabe erreicht. Bei allen Aufgaben, die diese Auffälligkeiten zeigen, ist daher zu diskutieren, ob sie im Aufgabenpool verbleiben können (vgl. Bühner, 2010; Rost, 2004).

Bis auf die ungeordneten Schwellenparameter der Aufgaben T3, Ä1, Ä2 und Ä3 lassen sich alle anderen Auffälligkeiten dieser Aufgaben und der Aufgaben EM1, EM2 und A2 durch die geringen Häufigkeiten der Antwortkategorie 3 (Niveau III) erklären. Die Häufigkeiten lagen entweder unter fünf Prozent oder unter zehn Prozent (vgl. Tab. 12). Somit liegen diese Auffälligkeiten nicht an der Qualität der Aufgaben und können daher auch nicht aus dem Aufgabenpool ausgeschlossen werden. Vielmehr lässt sich aus psychometrischer Sicht diskutieren, ob die gewählte Graduierung sinnvoll ist, weil nur eine geringe Anzahl von Personen die Antwortkategorie 3 (Niveau III) bedienen kann und damit die Messgenauigkeit reduziert wird (vgl. Rost, 2004; Kauertz, 2014). Demzufolge wäre zu überlegen, ob für die Teilkompetenzen *Eigenschaften von Modellen*, *Alternative Modelle*, *Testen von Modellen* und *Ändern von Modellen* das Scoring verändert werden sollte. Rost (2004) schlägt vor, aufeinanderfolgende Niveaus (hier Niveau II und III) zusammenzulegen. Bei der Teilkompetenz *Zweck von Modellen* besteht diese Problematik nicht, weil die prozentuale Verteilung über die Niveaus hinweg ausgeglichen ist.

Aus fachdidaktischer Sicht ist diese aus drei Niveaus bestehende Feinstruktur jedoch sinnvoll und sollte beibehalten werden, weil sie eine differenzierte Sicht auf die Lernenden ermöglicht und damit etwa wertvolle Informationen für Rückmeldungen an die Lernenden liefern kann. Daher wäre eine andere Möglichkeit in den Blick zu nehmen: Theoriegeleitet hängt dieser Befund damit zusammen, dass sowohl die methodische als auch die mediale Sichtweise auf Modelle jeweils als ein Kontinuum zu verstehen ist (Mahr, 2009; Oh & Oh, 2011). Dementsprechend ist es jedoch möglich, dass Kategorien, die derzeit noch dem Niveau II zugeordnet sind, so stark eine methodische Perspektive auf Modelle implizit einbeziehen, dass sie möglicherweise dem Niveau III zu-

geordnet werden könnten. Hier ist, gegebenenfalls unter Rückgriff auf das in dieser Studie entwickelte Kategoriensystem (vgl. Grünkorn, Upmeyer zu Belzen et al., 2014), das die einzelnen Niveaus näher ausdifferenziert, zu prüfen, wie die theoretische Abgrenzung von Niveau II und III anhand der empirischen Datenlage weiter geschärft werden kann.

Wie bereits ausgeführt, lassen sich die ungeordneten Schwellenparameter zwischen dem ersten und zweiten Übergang ($\delta_1 > \delta_2$) bei den Aufgaben T3, Ä1, Ä2 und Ä3 nicht mit einem geringen Vorkommen der Antwortkategorien erklären. Hier lagen die Häufigkeiten für die Antwortkategorien 1 (Niveau I) und 2 (Niveau II) über zehn Prozent (vgl. Tab. 12). Mögliche Gründe hierfür könnten die Aufgabenqualität oder die gewählte Graduierung der jeweiligen Teilkompetenz sein. Bei der Teilkompetenz **Testen von Modellen** wies nur die Aufgabe **Käfer** (T3) ungeordnete Schwellenparameter (Item-Deltas) auf, was für eine geringe Aufgabenqualität spräche. Demzufolge wäre zu empfehlen, die Aufgabe **Käfer** (T3) aus dem Aufgabenpool auszuschließen. Bei der Teilkompetenz **Ändern von Modellen** betrifft diese Auffälligkeit alle drei Aufgaben (Ä1, Ä2 und Ä3). Dieser Befund spricht daher zunächst nicht für einen Ausschluss der Aufgaben.

Eine weitere Möglichkeit wäre, die Graduierung beider Teilkompetenzen zu hinterfragen. Aufgrund der ungeordneten Schwellenparameter ($\delta_1 > \delta_2$) kann angenommen werden, dass die geforderten Kompetenzen auf Niveau I schwieriger sind als die auf Niveau II. Zudem zeigte die Wright Map (vgl. Abb. 12, S. 116), dass für das Erreichen von Niveau II der Teilkompetenzen **Testen von Modellen** und **Ändern von Modellen** nur geringe Fähigkeiten (< 0 Logits) benötigt werden. Unterstützt wird diese Annahme durch die Ergebnisse aus Studien von Krell (2012) und Terzer (2013), die im Rahmen von IRT-Analysen für die Niveaus Schwierigkeiten ermittelten. Die Ergebnisse zeigten, dass bei der Studie von Krell (2012) das Niveau II der Teilkompetenzen **Testen von Modellen** und **Ändern von Modellen** leichter war als Niveau I. Bei Terzer (2013) konnte dies nur für die Teilkompetenz **Testen von Modellen** nachgewiesen werden. Dies lässt sich mit unterschiedlichen Operationalisierungen von Niveau II der Teilkompetenz **Ändern von Modellen** erklären. Während sich bei Terzer (2013) die MC-Items des Niveaus II auf neue Erkenntnisse bezüglich des Originals beziehen, ist bei den Aufgaben im offenen Format die Kategorie **Ändern bei mangelnder Passung mit dem Original** zentral. Vor dem Hintergrund dieser Befunde wird empfohlen, entweder die entsprechenden Aufgaben auszuschlie-

Ben oder die vorgenommene Graduierung von Niveau I und II der Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* zu überdenken.

Zur Beurteilung des Tests ist außerdem die Varianz, die Reliabilität und die Abdeckung des Fähigkeitsspektrums anhand der Wright Map zu betrachten. Bei der eindimensionalen Skalierung lag die $Var < 1$ und deutete auf eine nicht ausreichende Abdeckung des Fähigkeitsspektrums hin (vgl. Neumann, 2014; Rost, 2004). Dies konnte durch die Wright Map bestätigt werden (vgl. Abb. 12).

Die Wright Map zeigte insgesamt, dass der Leistungstest den unteren Randbereich des Fähigkeitsspektrums (< 0) gut abdeckt. Somit können Personen, die Fähigkeiten in diesem Bereich haben, besonders gut unterschieden werden. Auffällig ist jedoch die Lücke in der Mitte des Fähigkeitsspektrums und im oberen Fähigkeitsbereich. Bislang liegen die Antwortkategorien einiger Aufgaben, zum Beispiel Ä1.3 (Antwortkategorie 3 der Aufgabe *Libelle* der Teilkompetenz *Ändern von Modellen*), über dem oberen Fähigkeitsbereich (vgl. Abb. 12). Das bedeutet, dass Personen mit hohen Fähigkeiten diese Antwortkategorien nicht erreichen können. Somit lassen sich über diese Personen nur schwer Aussagen treffen. Die Lücke im mittleren Fähigkeitsbereich ist allerdings noch bedeutender. In diesem Bereich ist ein Großteil der Personen lokalisiert. Auf dieser Grundlage können die Personenparameter nur schlecht geschätzt werden.

Daraus lässt sich schließen, dass für eine Verwendung des Tests im probabilistischen Ansatz weitere Aufgaben im offenen Antwortformat für den mittleren Fähigkeitsbereich zu konstruieren sind oder dass dieser Test mit anderen Tests zur Modellkompetenz, wie denen von Krell und Krüger (2010; Krell, Czeskleba et al., 2012) oder Terzer (2013), kombiniert werden müsste. In den extremen Randbereichen deckt der Test das Spektrum gut ab, deshalb können im Rahmen von Veränderungsmessungen in diesem Bereich differenzierte Aussagen getätigt werden. Somit ist davon auszugehen, dass dieser Test bei einer gemeinsamen Skalierung mit den MC-Items von Terzer (2013) die Lücke im Fähigkeitsspektrum schließen kann, die Patzke et al. (*in Druck*) in ihrem Beitrag zur Entwicklung von Modellkompetenz im Längsschnitt konstatieren.

Neben der Varianz ist auch die EAP/PV-Reliabilität des Tests, die durch die geringe Varianz beeinflusst wird, als schlecht zu interpretieren (Rost, 2004). Der Grund hierfür ist, neben der geringen Varianz, in der gewählten Testlänge zu suchen. Bei einer geringen Varianz kann nicht mit einer hohen Reliabilität ge-

rechnet werden, weil die Varianz eng mit der Reliabilität zusammenhängt (z. B. Amelang & Zielinski, 2002). Weiterhin ist zu berücksichtigen, dass jede Person nur drei Aufgaben bearbeitet hat und pro Teilkompetenz aufgrund des MMX-Designs entsprechend weniger. Mit einer Erhöhung der Anzahl der Aufgaben pro Testheft kann eine höhere Reliabilität des Tests erreicht werden. Wird beispielsweise eine eindimensionale Skalierung angestrebt, könnte nach der Spearman-Brown-Korrektur eine Testlänge von neun Aufgaben zu einer EAP/PV-Reliabilität von .72 führen (Rost, 2004). Diese Anzahl von Testaufgaben im offenen Antwortformat ist allerdings nach den Erfahrungen in den Teilstudien 1 und 2 nicht zu empfehlen. Mit längeren Bearbeitungszeiten sinkt die Bearbeitungsqualität, sodass diese nicht zu zuverlässigen Ergebnissen führen (*Optimizing-Satisficing-Modell*, Krosnick, 1999, zitiert nach Jonkisz et al., 2012).

5.4 Fazit

Mit der Umsetzung des zweiten Ziels dieser Forschungsarbeit – der Analyse und Beschreibung von Modellkompetenz – liegt sowohl ein inhaltlich überprüfbares als auch durch die formulierten Kategorien inhaltlich ausdifferenziertes Kompetenzmodell der Modellkompetenz vor. Hinsichtlich der Teilkompetenzen *Alternative Modelle*, *Testen von Modellen* und *Ändern von Modellen* wird empfohlen, ein basales Niveau zu berücksichtigen. In den Teilkompetenzen *Eigenschaften von Modellen* und *Zweck von Modellen* werden die Perspektiven der Lernenden umfassend beschrieben.

Ausgehend von der qualitativen Inhaltsanalyse der Besonderheiten in den Antwortstrukturen der Lernenden, konnte ein vertieftes Verständnis der Sichtweisen der Lernenden auf Modelle entwickelt werden. So konnten bei den Schülerinnen und Schülern dieser Stichprobe didaktisch orientierte Perspektiven (*Verständlichkeit/Kommunizierbarkeit* und *Zugänglichkeit*) beschrieben werden, die sich über alle fünf Teilkompetenzen und eine mediale sowie methodische Perspektive (bzw. Niveau) erstrecken.

Weiterhin legen die Befunde nahe, bei Lernenden zwischen einem *abstrakten* und einem *konkreten Modellverständnis* zu unterscheiden. Damit unterstützen diese Befunde die Ergebnisse anderer Studien (vgl. Krell, Upmeyer zu Belzen et al., 2012b; Terzer, 2013), die ebenfalls eine Unterscheidung zwischen diesen vornehmen und darüber hinaus bei Lernenden die Schwierigkeit erkennen,

abstrakte Vorstellungen auf konkrete Vorstellungen von Modellen zu übertragen.

Die Analysen zeigten überdies, dass das Parallelisieren von Modell und Original für die Schülerinnen und Schüler in den Teilkompetenzen **Testen** und **Ändern von Modellen** zentral ist. Auffällig ist, dass Lernende zum Teil zweckungebunden parallelisieren. Das vorherrschende Parallelisieren ist unter den Schülerinnen und Schülern offensichtlich ein grundlegender und bewusst zu berücksichtigender Aspekt bei der Arbeit mit Modellen (vgl. S. 137ff. dieser Arbeit).

In Bezug auf den Einsatz des entwickelten Leistungstests in einem probabilistischen Ansatz lässt sich festhalten, dass dieser zwar vertretbar ist, dass es jedoch notwendig ist, weitere Aufgaben zu entwickeln oder den Test mit anderen Tests zur Modellkompetenz zu kombinieren, um das Fähigkeitsspektrum besser abzudecken (vgl. Patzke et al., *in Druck*). Die Befunde der Analyse weisen zudem darauf hin, dass das Scoring (Niveau I, II und III) insbesondere in Bezug auf die Teilkompetenzen **Testen von Modellen** und **Ändern von Modellen** überdacht werden muss.

6 Didaktische Ableitungen und Anregungen

Nach der Einführung der Bildungsstandards stehen Lehrkräfte vor der Herausforderung, eine an Kompetenzen ausgerichtete Lernkultur aufzubauen (KMK, 2010). Ein solcher Unterricht ist ein **schüleraktiver Unterricht**, in dem „... keine ‚trägen‘ und isolierten Kenntnisse und Fähigkeiten entstehen, sondern anwendungsfähiges Wissen und ganzheitliches Können, das zum Beispiel reflektive und selbstregulative Prozesse einschließt“ (Klieme & Hartig, 2007, S. 3). In diesem Sinne haben Biologie-Lehrkräfte die Aufgabe, Modellkompetenz bei Schülerinnen und Schülern zu entwickeln. Dies bedeutet, dass sie die Lernenden darin unterstützen, Modelle sowohl als Mittel der Veranschaulichung als auch in ihrer Rolle im wissenschaftlichen Prozess als Mittel der Erkenntnisgewinnung wahrzunehmen. Um diese beiden Aufgaben erfolgreich zu bewältigen, benötigen Lehrkräfte Evidenz aus empirischen Analysen, alltagstaugliche Instrumente sowie konkrete didaktische Anregungen (Abb. 13).

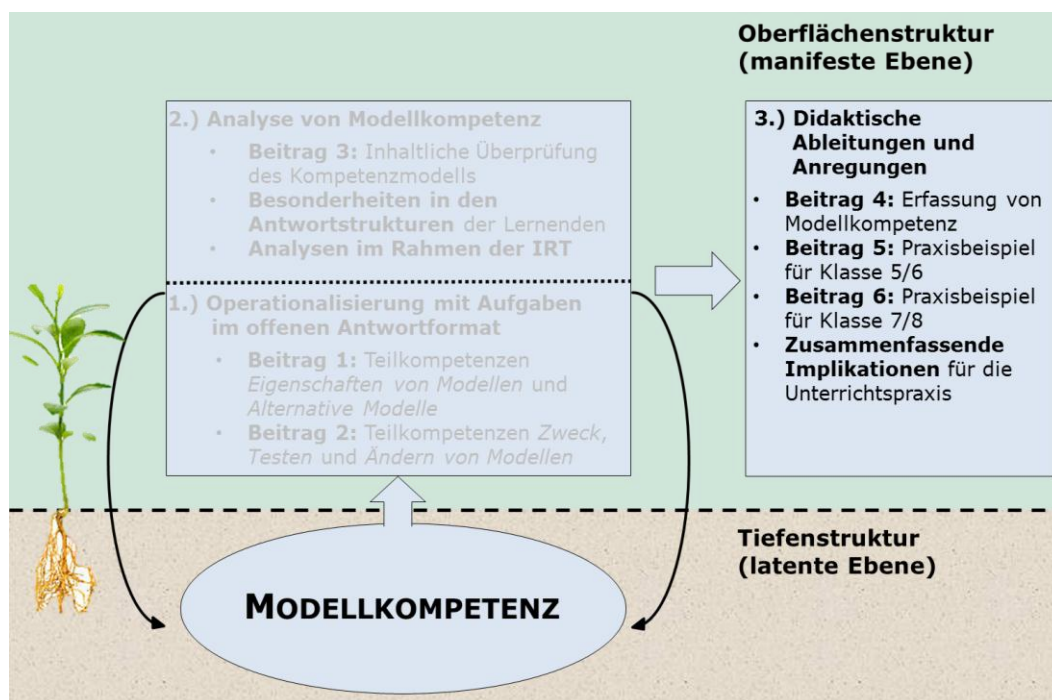


Abb. 13: Drittes Ziel der Forschungsarbeit: Didaktische Ableitungen und Anregungen aus den Analysen (Grafik nach Meyer, 2012)

In der Literatur wurden bereits Hinweise auf eine Förderung der Modellkompetenz im Unterricht gegeben (u. a. Fleige et al., 2012a, 2012b; Füssenich et al., 2013; Grosslight et al., 1991; Justi & Gilbert, 2003; Leisner-Bodenthin, 2006;

Louca et al., 2011; Schwarz et al., 2009). Ergänzend dazu werden auf der Basis der Ergebnisse der vorliegenden Arbeit alltagstaugliche Instrumente zur Erfassung von Modellkompetenz (Grünkorn, Lotz et al., 2014) und konkrete Anregungen für den Unterricht abgeleitet (Grünkorn & Fleige, 2012; Grünkorn & Hanauer, 2013a, 2013b; vgl. S. 137ff. dieser Arbeit).

6.1 Eigene Beiträge zu didaktischen Anregungen

Die didaktischen Empfehlungen und Anregungen zur Förderung von Modellkompetenz im Biologieunterricht sind in drei Publikationen veröffentlicht:

- Grünkorn, J., Lotz, A. & Terzer, E. (2014). Erfassung von Modellkompetenz im Biologieunterricht. *Der mathematische und naturwissenschaftliche Unterricht*, **67** (3), 132-138.
- Grünkorn, J. & Hanauer, N. (2013b). Schwimmen wie Fische im Wasser – Untersuchung der Schleimschicht von Fischen. *Grundschule*, **45** (6), 28-29.
Grünkorn, J. & Hanauer, N. (2013a). Fischhaut [Materialien]. *Grundschule*, **45** (6), Beilage, XIII-XVI.
- Grünkorn, J. & Fleige, J. (2012). Bau und Funktion der Fischhaut. In J. Fleige, A. Seegers, A. Upmeyer zu Belzen & D. Krüger (Hrsg.), *Modellkompetenz im Biologieunterricht Klasse 7-10. Phänomene begreifbar machen – in 11 komplett ausgearbeiteten Unterrichtseinheiten* (S. 23-28). Donauwörth: Auer.

Die in den Publikationen beschriebenen Empfehlungen und Anregungen greifen zum Teil die von Fleige et al. (2012a, 2012b) gegebenen Hinweise zur Förderung von Modellkompetenz auf. Dort werden die Reflexion über Modelle und ihre Teilkompetenzen, die Nutzung einfacher Fachinhalte mit einem Fokus auf die Förderung von Modellkompetenz, der Fokus auf einzelne Teilkompetenzen und die Formulierung von Merksätzen aufgeführt.

6.1.1 Beitrag 4: Erfassung von Modellkompetenz im Biologieunterricht

Grünkorn, J., Lotz, A. & Terzer, E. (2014). Erfassung von Modellkompetenz im Biologieunterricht. *Der mathematische und naturwissenschaftliche Unterricht*, 67 (3), 132-138.

Zusammenfassung

Für ein effektives kompetenzorientiertes Unterrichten ist eine lernbegleitende Erfassung des Kompetenzstands von Lernenden förderlich. Eine solche Erfassung informiert über die bereits vorhandenen Kompetenzen der Lernenden und macht deutlich, welche Kompetenzen noch erworben werden müssen. Diese Informationen können den Lernprozess unterstützen, wenn Schülerinnen und Schüler im Sinne eines formativen Assessments Rückmeldungen über ihren Kompetenzstand erhalten und dieser bei der Unterrichtsplanung berücksichtigt wird (u. a. Black & William, 2009; Hattie & Timperley, 2007).

Der Beitrag präsentiert daher einen Diagnosebogen und ein Kompetenzraster, das für die Erfassung von Modellkompetenz herangezogen werden kann. Hierauf aufbauend können Lehrkräfte adäquate, individuell angepasste Fördermaßnahmen entwickeln und anbieten.

Für den Diagnosebogen vervollständigen die Lernenden vorgegebene Satzanfänge. Diese Satzanfänge sind in Anlehnung an Interviewfragen von Grosslight et al. (1991) formuliert. Für die Teilkompetenz *Eigenschaften von Modellen* lautet der Satzanfang beispielsweise „Die Beziehung zwischen Modell und Original besteht darin, dass ...“ (Grünkorn, Lotz et al., 2014, S. 136). Lehrkräfte ordnen die Antworten der Lernenden mithilfe eines Kategoriensystems zur Modellkompetenz ein und können damit den Kompetenzstand des Lernenden einschätzen (vgl. Kunter & Klusmann, 2010). Das dafür verwendete Kategoriensystem beschreibt verschiedene Perspektiven von Schülerinnen und Schülern auf Modelle und auf den Modellbildungsprozess, mit denen Lehrkräfte im Unterricht rechnen können (vgl. Grünkorn, Upmeyer zu Belzen et al., 2014).

Darüber hinaus wird ein erprobtes Kompetenzraster vorgestellt, das Schülerinnen und Schülern lernbegleitend Rückmeldungen zu ihrem Kompetenzstand gibt und sie auf diese Weise dabei unterstützt, ihren Lernprozess zielgerichtet zu strukturieren. Lernende können dieses Kompetenzraster nutzen, um ihre Kompetenzen selbst einzuschätzen. Somit können subjektive Sichtweisen bzw.

Selbstkonzepte der Lernenden dokumentiert werden (vgl. Kunter & Klusmann, 2010). Das Kompetenzraster umfasst verschiedene „Ich kann“-Beschreibungen, die auf dem veröffentlichten Kategoriensystem (Grünkorn, Upmeier zu Belzen et al., 2014), der Operationalisierung von Modellkompetenz in MC-Items (Terzer et al., 2013) sowie auf Vorstellungen der Lernenden, die diese während der Bearbeitung von MC-Items mit der Methode des lauten Denkens geäußert haben (Terzer et al., 2012), beruhen. Für das Niveau I der Teilkompetenz *Eigenschaften von Modellen* lautet die „Ich kann“-Beschreibung beispielsweise: „Ich kann Eigenschaften beschreiben, in denen Modell und Original sich gleichen“ (Grünkorn, Lotz et al., 2014, S. 137). Das Kompetenzraster zeigt abstrakt gehaltene Formulierungen, die mit Bezug auf konkrete Kontexte aus dem jeweiligen Unterricht umformuliert werden können. Die Nutzung beider Varianten – abstrakte und konkrete Formulierungen – ist zu empfehlen, um das abstrakte wie das konkrete Modellverständnis der Lernenden zu fördern (vgl. S. 120ff. dieser Arbeit; Grünkorn, Upmeier zu Belzen et al., 2014). Die Erfassung von Modellkompetenz mithilfe der beiden Instrumente sollte zu Beginn und während der Unterrichtseinheit erfolgen. Um dies möglichst effektiv zu gestalten, sind Phasen der Erfassung und Phasen der Förderung von Kompetenzen zyklisch miteinander verbunden (vgl. Schröder & Wirth, 2012). In einer zusammenfassenden Schlussreflexion nach Unterrichtseinheiten zu Modellen kann erneut auf das konkrete und das abstrakte Modellverständnis eingegangen werden.

6.1.2 Beitrag 5: Schwimmen wie Fische im Wasser – Untersuchung der Schleimschicht von Fischen

Grünkorn, J. & Hanauer, N. (2013b). Schwimmen wie Fische im Wasser – Untersuchung der Schleimschicht von Fischen. *Grundschule*, 45 (6), 28-29.

Grünkorn, J. & Hanauer, N. (2013a). Fischhaut [Materialien]. *Grundschule*, 45 (6), Beilage, XIII-XVI).

Zusammenfassung

Das Praxisbeispiel zum Thema *Angepasstheit von Lebewesen an ihren Lebensraum am Beispiel von Fischen* zielt auf die Förderung von Lernenden der Jahrgangsstufe 5/6 in den Teilkompetenzen *Zweck von Modellen* und *Testen von Modellen*. Die Fokussierung auf einzelne Teilkompetenzen bei der Förderung von Modellkompetenz wird von Fleige et al. (2012a, 2012b) empfohlen. Die Kombination dieser Teilkompetenzen wurde gewählt, um insbesondere das zweckgebundene *Testen von Modellen* zu fördern (vgl. S. 120ff. dieser Arbeit; Grünkorn, Upmeyer zu Belzen et al., 2014).

Ausgehend von einem Phänomen, das für Lernende einfach verständlich und interessant ist (vgl. Fleige et al., 2012a, 2012b), werden eine Fragestellung zum Effekt der Schleimschicht auf die Fortbewegungseigenschaften von Fischen und eine zu prüfende Hypothese herausgearbeitet. Zur Überprüfung dieser Hypothese bauen Schülerinnen und Schüler Modelle und setzen diese in einem Versuch ein, um Erkenntnisse über das Original zu gewinnen. Damit Lernende die Modelle methodisch nutzen, ist es empfehlenswert, sie diese Modelle selbst bauen zu lassen (vgl. Fleige et al., 2012a).

Um ausreichend viele Gelegenheiten zur Reflexion zu schaffen, bietet sich an verschiedenen Stellen im Unterrichtsgang der Einsatz der folgenden Reflexionsimpulse an (vgl. Formulierungen der Fragestellungen für die Teilkompetenzen, Grünkorn & Krüger, 2012, S. 16):

- *Zweck von Modellen:* Gib an, welchen Zweck Modelle haben.
- *Testen von Modellen:* Erkläre, wie wir überprüfen können, ob die Modelle tauglich sind und ihren Zweck erfüllen.

Beispielsweise kann die Lehrkraft bereits zur Ableitung der Hypothese selbstgebaute Modelle einsetzen und an dieser Stelle über den *Zweck dieser Modelle* – die Voraussagekraft – sprechen. Dies kann sie vor Beginn und am Ende des eigentlichen Versuchs zur Überprüfung der Hypothese wiederholen. Zudem sollte auch hier über das *Testen von Modellen* reflektiert werden.

In der abschließenden Reflexionsphase ist in Form von Merksätzen (vgl. Fleige et al., 2012b) Folgendes zu sichern: Modelle dienen dazu, Vorhersagen zu treffen bzw. Hypothesen abzuleiten und zu untersuchen (*Zweck von Modellen*). Erst durch die Anwendung der Modelle und den Gewinn von Daten aus dem

Modellversuch können die Hypothesen überprüft werden (*Testen von Modellen*).

Neben diesen Reflexionsphasen kann die Lehrkraft spezifisch auf die Förderung der Teilkompetenz *Testen von Modellen* eingehen, wenn ein unsystematisches Vorgehen oder eine gewisse Ratlosigkeit bei den Lernenden zu beobachtet ist. Hierzu eignet sich eine Schritt-für-Schritt-Prozedur (vgl. S. 120ff. dieser Arbeit). Es werden drei Schritte vorgeschlagen, die sowohl die mediale Perspektive (Schritte 1 und 2) als auch die methodische Perspektive (Schritt 3) in den Blick nehmen (Füssenich et al., 2012, S. 11). Wichtig ist, dass diese Prozedur mit Bezug auf den Zweck des Modells im Unterricht eingesetzt wird.

6.1.3 Beitrag 6: Bau und Funktion der Fischhaut

Grünkorn, J. & Fleige, J. (2012). Bau und Funktion der Fischhaut. In J. Fleige, A. Seegers, A. Upmeier zu Belzen & D. Krüger (Hrsg.), *Modellkompetenz im Biologieunterricht Klasse 7-10. Phänomene begreifbar machen – in 11 komplett ausgearbeiteten Unterrichtseinheiten* (S. 23-28). Donauwörth: Auer.

Zusammenfassung

Im Unterschied zum oben referierten Beitrag 5 schlägt das Unterrichtskonzept in Beitrag 6 zum Thema *Bau und Funktion der Fischhaut* für Lernenden der Jahrgangsstufe 7/8 Fördermaßnahmen für alle fünf Teilkompetenzen des Kompetenzmodells vor. Es ist der Lehrkraft überlassen, welchen Fokus sie im Unterricht setzt.

Auch in dieser Unterrichtseinheit werden, ausgehend von einem Phänomen, eine Fragestellung zum Effekt der Anordnung der Fischschuppen auf die Eigenschaften der Fischhaut und eine entsprechende Hypothesen entwickelt, dazu Modelle erstellt und mit ihnen die Hypothesen in der Anwendung des Modells geprüft. Auf der Basis der Ergebnisse werden die Modelle gegebenenfalls geändert.

Im Mittelpunkt dieses Beitrags steht die Förderung eines konkreten und abstrakten Modellverständnisses (vgl. S. 120ff. dieser Arbeit; Grünkorn, Upmeier zu Belzen et al., 2014). Hierfür kann das Schema zum Einsatz von Modellen zur Erkenntnisgewinnung (vgl. Fleige et al., 2012b) genutzt werden. Mit diesem Schema kann dieser Prozess zunächst auf konkreter Ebene reflektiert werden. Im Anschluss daran können die Schülerinnen und Schüler mithilfe des

Schemas von diesem konkreten Inhalt abstrahieren und damit ein abstraktes Modellverständnis entwickeln.

Im Laufe der Entwicklung und Anwendung dieser Modelle sowie am Ende der Unterrichtseinheit sind die einzelnen Teilkompetenzen zu reflektieren (vgl. Fleige et al., 2012a, 2012b). Die Stellen, an denen die Reflexion erfolgen soll, sind in einem Schema gekennzeichnet (Grünkorn & Fleige, 2012, S. 28). Zur Unterstützung dieses Reflexionsprozesses sind für die jeweilige Teilkompetenz ein Impuls und Antwortmöglichkeiten auf allen drei Niveaus gegeben (vgl. Formulierungen der Fragestellungen für die Teilkompetenzen, Grünkorn & Krüger, 2012, S. 16; Grünkorn et al., 2011, S. 56f.). Die Antwortmöglichkeiten orientieren sich an den Antworten der Lernenden aus den Studien der vorliegenden Arbeit und können von der Lehrkraft als Orientierungshilfe genutzt werden.

6.2 Zusammenfassende Implikationen für die Unterrichtspraxis

Auf der Basis der Ergebnisse aus den Studien zur Analyse von Modellkompetenz bei Schülerinnen und Schülern der Sekundarstufe I wurde das dritte Ziel dieser Arbeit – die Ableitung didaktischer Anregungen und Empfehlungen – realisiert. Die abgeleiteten Empfehlungen unterstützen und ergänzen die bereits erwähnten Hinweise auf eine Förderung von Modellkompetenz von Fleige et al. (2012a, 2012b).

Die Befunde aus der vorliegenden Arbeit informieren über verschiedene Perspektiven von Schülerinnen und Schülern auf Modelle und den Modellbildungsprozess (Grünkorn, Upmeier zu Belzen et al., 2014) und legen nahe, sowohl ein konkretes als auch ein abstraktes Modellverständnis in den Unterricht einzubinden (vgl. S. 120ff. dieser Arbeit; Grünkorn, Upmeier zu Belzen et al., 2014), die Begründungen für Perspektiven auf die Teilkompetenz *Eigenschaften von Modellen* in den Unterricht einzubeziehen (Grünkorn, Upmeier zu Belzen et al., 2014), die Zweckgebundenheit von Modellen zu thematisieren (vgl. S. 120ff. dieser Arbeit; Grünkorn, Upmeier zu Belzen et al., 2014) und die Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* in einer Schritt-für-Schritt-Prozedur zu fördern (vgl. S. 120ff. dieser Arbeit).

Erkenntnisse hinsichtlich der Perspektiven von Lernenden

Insgesamt ergibt sich aus den Ergebnissen dieser Forschungsarbeit das Bild, dass Lernende über eine Vielfalt von unterschiedlichen Perspektiven auf Modelle und auf den Modellbildungsprozess verfügen (Grünkorn, Upmeier zu Belzen et al., 2014). Darunter sind beispielsweise auch basale Perspektiven für die Teilkompetenzen *Alternative Modelle*, *Testen von Modellen* und *Ändern von Modellen* (Grünkorn, Upmeier zu Belzen et al., 2014) subsumiert. Diese Perspektiven von Schülerinnen und Schülern zeichnen sich durch eine Ablehnung grundlegender Konzepte aus, die für eine Entwicklung der jeweiligen Teilkompetenz allerdings wesentlich sind. Daher ist es wichtig, diese Perspektiven zu kennen, um zu den Lernenden individuell passende Fördermaßnahmen zu entwickeln (vgl. Grünkorn, Lotz et al., 2014).

Die Ergebnisse der Teilkompetenz *Alternative Modelle* legen dabei nahe, dass es nicht ausreicht, lediglich verschiedene Modelle zu einem Original im Unterricht zu präsentieren (Grünkorn, Upmeier zu Belzen et al., 2014). Stattdessen sollten Modelle verstärkt diskutiert und reflektiert werden (vgl. Fleige et al., 2012a, 2012b). Zudem deuten die Ergebnisse darauf hin, dass ein reflektierter Umgang mit historischen Modellen (z. B. Biomembranmodelle) im Unterricht wichtig ist (Grünkorn, Upmeier zu Belzen et al., 2014). Dies fordern auch Justi und Gilbert (2002) auf der Basis ihrer Studie mit Lehrkräften. Erfolgt ein reflektierter Umgang mit historischen Modellen, ist anzunehmen, dass die Lernenden elaborierte Vorstellungen von Modellen hinsichtlich von Modellen als Mittel der Erkenntnisgewinnung ausbilden. Lernende erkennen, dass Forschende über längere Zeit hinweg verschiedene Hypothesen über ein Original entwickeln und dass diese sich stets verändern können. Erfolgt kein reflektierter Umgang mit historischen Modellen im Unterricht, können die Lernenden den Eindruck gewinnen, dass Modelle endgültig sind, sobald sie einen längeren Prozess durchlaufen haben und in Büchern dokumentiert sind.

Basale Perspektiven in den Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* zeichnen sich durch eine Ablehnung des Testens und Änderns von Modellen aus (Grünkorn, Upmeier zu Belzen et al., 2014). Um Kompetenzen auf allen drei Niveaus der beiden Teilkompetenzen bei den Lernenden aufzubauen, kann es für die Schülerinnen und Schüler hilfreich sein, den Modellbildungsprozess selbst zu erfahren, d. h. Modelle selbst zu bauen und einzusetzen (vgl. Fleige et al., 2012a, 2012b). So kann ihnen die Notwendigkeit ei-

nes Testens von Modellen und die damit verbundene Änderung von Modellen verdeutlicht werden (vgl. Grünkorn & Fleige, 2012; Grünkorn & Hanauer, 2013b; Grünkorn, Lotz et al., 2014).

Damit Lehrkräfte angepasste Fördermaßnahmen entwickeln können, ist es wichtig, dass sie den Kompetenzstand der eigenen Lerngruppe diagnostizieren. Instrumente, die für eine solche Individualdiagnostik geeignet sind, liegen derzeit nicht vor. Gegenwärtig arbeiten Gogolin und Krüger (*in Druck*) auf der Basis des hier entwickelten Kategoriensystems an einem solchen Instrument. Für die Einschätzung des Kompetenzstands von Schülerinnen und Schülern eignen sich über individualdiagnostische Tests hinaus Kompetenzraster oder Diagnosebögen, die im Beitrag von Grünkorn, Lotz et al. (2014) präsentiert werden.

In Bezug auf eine adäquate Einschätzung der Modellkompetenz von Lernenden konstatieren Fleige et al. (2012b), dass angehende Lehrkräfte nicht hinreichend über eine entsprechende Diagnosekompetenz verfügen. Daher untersuchen Günther, Fleige, Upmeyer zu Belzen und Krüger (2014) in einem Projekt zur Förderung professioneller Kompetenzen von Lehramtsstudierenden, ob eine Förderung der Diagnose- und Vermittlungskompetenz mit fallbasierten Lernaufgaben gelingt. Ergänzend dazu lässt sich – wie in Beitrag 5 vorgeschlagen wird – das Kategoriensystem heranziehen, das unter anderem Beispielaussagen von Schülerinnen und Schülern pro Kategorie anführt. Diese authentischen Aussagen ordnen Lehrkräfte dem Kategoriensystem bzw. dem Kompetenzmodell zu, und dadurch lässt sich ihre Diagnosekompetenz trainieren (vgl. Fleige et al., 2012b).

Weiterhin kann mit dem Kategoriensystem im Rahmen von Fortbildungen auch die Modellkompetenz von angehenden und erfahrenen Lehrkräften weiterentwickelt werden, indem ihnen die Perspektivenvielfalt (im Sinne einer medialen und methodischen Sichtweise) auf Modelle bewusst gemacht wird. Die Relevanz einer solchen Förderung machen empirische Studien deutlich, die zeigten, dass Lehrkräfte häufig über eine Modellkompetenz verfügen, die mit denen von Schülerinnen und Schülern vergleichbar ist (u. a. Crawford & Cullin, 2004, 2005; Fleige et al., 2012b; Justi & Gilbert, 2002, 2003; van Driel & Verloop, 1999, 2002).

Förderung eines konkreten und abstrakten Modellverständnisses

In den Teilkompetenzen *Zweck von Modellen* und *Testen von Modellen* zeigten sich bei den Antworten der Lernenden wiederholt Unterschiede zwischen abstrakten und konkreten Vorstellungen von Modelle und der Modellbildung (Grünkorn, Upmeier zu Belzen et al., 2014; Krell, Upmeier zu Belzen et al., 2012b; Leisner & Mikelskis, 2004; Terzer, 2013). Da ein abstraktes und ein konkretes Modellverständnis Teil einer umfassenden Modellkompetenz sind, ist es wichtig, beide Perspektiven im Unterricht in den Blick zu nehmen. Hierfür eignet sich das in den Beiträgen 4, 5 und 6 (Grünkorn & Fleige, 2012; Grünkorn & Hanauer, 2013b; Grünkorn, Lotz et al., 2014) bereits angesprochene Schema von Fleige et al. (2012b) für den Einsatz von Modellen zur Erkenntnisgewinnung, das auf der Basis der „*components of a scientific episode*“ von Giere et al. (2006, S. 36) entwickelt wurde.

Einbindung von Reflexionsphasen

Insgesamt zeigen die Ergebnisse, dass Schülerinnen und Schüler im Gegensatz zu Forschenden Modelle vorwiegend als Medien und weniger als Mittel der Erkenntnisgewinnung verstehen (Grünkorn, Upmeier zu Belzen et al., 2014). Zudem wies die Wright Map (vgl. Abb. 12, S. 116) für die Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* im Vergleich zu den anderen Teilkompetenzen eine große Lücke zwischen dem *Thurstonian Threshold* 2 und 3 auf. Daraus lässt sich ableiten, dass zwischen Niveau II und III ein größerer Unterschied zu verzeichnen ist als zwischen Niveau I und II. Auch Krell, Upmeier zu Belzen et al. (2012b) stellten dies in ihren Untersuchungen zur Teilkompetenz *Zweck von Modellen* fest: “[...] it is comparatively ‘harder’ to take the step from level II to III than to take the step from level I to II” (S. 21). Insgesamt bestätigt dies, dass zwischen Niveau II und Niveau III, also zwischen einer medialen und einer methodischen Ebene, ein größerer inhaltlicher Unterschied besteht als zwischen Niveau I und II, die beide der medialen Ebene zuzuordnen sind. Es ist daher anzunehmen, dass die Voraussetzungen (*Unzufriedenheit*, *Verständlichkeit*, *Plausibilität* und *Fruchtbarkeit*, vgl. Krüger, 2007) für einen *Conceptual Change* (Strike & Posner, 1992) von Niveau II auf Niveau III von den Lernenden schwieriger zu erreichen sind als für den von Niveau I auf Niveau II.

Es ist also wichtig, ausreichend viele Wiederholungen und viel Zeit für Reflexionen einzuplanen, wie bereits von Fleige et al. (2012a; 2012b) empfohlen. Für solche Reflexionsphasen eignen sich Impulse, die in den Artikeln von Grünkorn et al. (2011) und Grünkorn und Krüger (2012) für die einzelnen Teilkompetenzen des Kompetenzmodells vorgeschlagen werden. Beitrag 5 (Grünkorn & Hanner, 2013b) und Beitrag 6 (Grünkorn & Fleige, 2012) geben für den Einsatz dieser Reflexionsimpulse Anregungen.

Eine weitere Möglichkeit, die methodische Perspektive auf Modelle zu entwickeln, könnte in der Reflexion der didaktisch orientierten Perspektiven liegen: **Modell als Mittel der Zugänglichkeit** und **Modell als Mittel der Verständlichkeit/Kommunizierbarkeit**. Die Ergebnisse haben gezeigt, dass Schülerinnen und Schüler in ihren Aussagen diese Perspektiven mit einer medialen und methodischen Sicht auf Modelle verknüpften (vgl. S. 120ff. dieser Arbeit). Damit können diese Perspektiven als Türöffner für ein elaboriertes Modellverständnis fungieren, indem diese Perspektiven im Unterricht explizit reflektiert werden. So könnte beispielsweise im Rahmen der Diskussion über den **Zweck von Modellen** die Verständlichkeit/Kommunizierbarkeit von Modellen thematisiert werden. Wichtig wäre es hier, herauszuarbeiten, dass Modelle zum einen dazu da sind, Wissen besser zu verstehen und zu vermitteln, und zum anderen ein wichtiges Kommunikationsmittel darstellen, um Ideen oder Annahmen untereinander auszutauschen und diese zu diskutieren. Während Schülerinnen und Schüler die erste Perspektive häufig einnehmen, erwähnen sie die zweite, komplexere Perspektive selten (vgl. S. 117ff. dieser Arbeit).

Neben den Reflexionsimpulsen und den didaktisch orientierten Perspektiven ist es empfehlenswert, den Modellbegriff im Unterricht zu besprechen. Dies erscheint deswegen relevant, weil in dieser Forschungsarbeit alternative Auffassungen vom Modellbegriff identifiziert wurden (Grünkorn et al., 2011). So stellen einige Schülerinnen und Schüler bei Zeichnungen oder Analogien keine Verbindung zu Modellen her. Hier ist anzunehmen, dass diese Auffassung von Modellen den Lernprozess bzw. den Vorstellungswandel behindert. Andere Lernende nehmen alles, was von einer Person (ihnen selbst oder anderen Personen) zum Modell erklärt wird, auch als Modell wahr. Diese Auffassung von Modellen ist elaboriert und greift das Modellverständnis dieser Arbeit auf (vgl. Mahr, 2003; Stachowiak 1973). Für die Entwicklung eines elaborierten Verständnisses des Modellbegriffs bei den Lernenden liefert Trier (2013) in ihrem

Unterrichtsbeispiel Anregungen für eine adäquate Auseinandersetzung mit dem Modellbegriff. Sie schlägt vor, verschiedene Formen von Modellen (z. B. Baupläne, echte Pflanzen, Formeln) im Unterricht einzuführen, diese explizit als Modelle zu benennen und sie als Modelle im Rahmen eines Experiments einzusetzen.

Anregungen zur Förderung der *Eigenschaften von Modellen*

Die Begründungen von Lernenden dafür, inwieweit ein Modell so aussieht wie das Original, könnten Lehrkräfte darin unterstützen, adäquate, angepasste Fördermaßnahmen zu entwickeln. Begründen die Lernenden beispielsweise ihre Sichtweise auf Modelle als Kopien der Originale mit einem großen Vertrauen in die Forschenden, dann ist es ratsam, den wissenschaftlichen Prozess, die Rolle der Forschenden sowie die Zugänglichkeit von Originalen noch stärker zu diskutieren (Grünkorn, Upmeier zu Belzen et al., 2014). Im Hinblick auf eine Förderung der Teilkompetenz *Eigenschaften von Modellen* ist es daher empfehlenswert, von Lernenden nicht nur die Beziehung zwischen Modell und Original beschreiben zu lassen, sondern zusätzlich nach Gründen für ihren Standpunkt zu fragen. Dies ermöglicht Einblicke in das Wissenschaftsverständnis (Mayer, 2007) von Lernenden und in ihre epistemologischen und ontologischen Positionen (Günther, 2006).

In diesem Zusammenhang können auch die Grenzen und der hypothetische Charakter von Modellen in gegenwärtigen Lehrbüchern und Museen besprochen werden. Insbesondere die Perspektive, nach der Lernende *Modelle als Mittel der Zugänglichkeit* wahrnehmen, um Originale direkt zu erfahren, ist kritisch zu reflektieren. Hier ist es wichtig, die Datengrundlage für eine Modellierung zu diskutieren, um zu vermeiden, dass Lernende Modelle als Kopien verstehen.

Beachtung der Zweckgebundenheit von Modellen

Zur Förderung einer medialen und methodischen Perspektive auf Modelle und auf den Modellbildungsprozess ist zudem eine zweckgebundene Betrachtung von Modellen essenziell. Dies liegt darin begründet, dass die Modellierenden darüber entscheiden, welche Eigenschaften des Originals für das Modell relevant sind. Vor diesem Hintergrund ist ein Modell immer an einen bestimmten Zweck gebunden (Bailer-Jones, 2002; Giere, 2004; Halloun, 2001; Mahr,

2009; Stachowiak, 1973). Somit ist bereits aus theoretischer Sicht zu empfehlen, den **Zweck von Modellen** gemeinsam mit anderen Teilkompetenzen zu fördern und eine Teilkompetenz nicht ohne Eingehen auf den Zweck des jeweiligen Modells zu thematisieren.

Auch die Ergebnisse dieser Forschungsarbeit sprechen dafür, dass beim Unterrichten der jeweiligen Teilkompetenzen bzw. bei allen Phasen der Modellierung (von der Entwicklung hin zur Beurteilung) der Zweck des jeweiligen Modells deutlich herausgestellt werden sollte (vgl. S. 120ff. dieser Arbeit; Grünkorn, Upmeier zu Belzen et al., 2014): Die Analysen zeigen, dass Lernende beispielsweise Modelle verändern, wenn sie in ihrer Struktur und/oder Funktion nicht mit dem Original übereinstimmen. An dieser Stelle zählt für den Lernenden eine genaue Übereinstimmung zwischen Original und Modell ungeachtet des Zwecks des Modells, für den es gebaut wurde.

Die zweckgebundene Konstruktion ist allerdings wichtig, um etwa beim **Testen von Modellen** oder **Ändern von Modellen** Qualitätskriterien für ein Modell formulieren zu können. Besonderes Augenmerk sollten die Lehrkräfte daher auf eine angemessene Beurteilung des Modells unter Berücksichtigung seines Zwecks legen, wie dies in Beitrag 5 (Grünkorn & Hanauer, 2013b) bereits exemplarisch gezeigt wurde. Das bedeutet, dass die im Unterricht durchgeführte Modellkritik mit ihren Kriterien der Entsprechung, der Verkürzung und des Beiwerks (Kattmann, 2006) zweckgebunden vorgenommen werden sollte.

Anregungen zur Förderung des Testens von Modellen und des Änderns von Modellen

Neben einem verstärkten Parallelisieren in den Teilkompetenzen **Testen von Modellen** und **Ändern von Modellen** konnten in beiden Teilkompetenzen weitere vergleichbare Kategorien identifiziert werden (Grünkorn, Upmeier zu Belzen et al., 2014). Dies lässt sich theoretisch begründen, denn das **Ändern von Modellen** ist eine Folge des **Testens**. Daher bietet es sich an, im Unterricht beide Teilkompetenzen kombiniert zu fördern (vgl. Grünkorn & Fleige, 2012).

Weiterhin deuten die Ergebnisse darauf hin, dass Lernende Schwierigkeiten haben, Gründe für ein **Ändern von Modellen** zu nennen (Grünkorn, Upmeier zu Belzen et al., 2014). Ihre Antworten sind zum Teil vage und weisen auf eine gewisse Ratlosigkeit hin. Einige Lernende zeigten überdies ein alternatives Verständnis des Begriffes **Ändern**. Es ist davon auszugehen, dass dies sie da-

ran hinderte, elaborierte Antworten zu geben (Grünkorn, Upmeier zu Belzen et al., 2014). So verbanden manche Schülerinnen und Schüler mit dem Begriff **Ändern** Änderungen des Modellzustands während der Benutzung des Modells (z. B. entspricht die Bewegung des Modells bereits einer Änderung). Zudem zeigte sich, dass in dieser Teilkompetenz im Vergleich zu anderen Teilkompetenzen nur wenige Schülerinnen und Schüler Niveau III erreichten.

Grosslight et al. (1991) stellten überdies fest, dass Lernende unstrukturiert antworteten, wenn sie nach dem **Testen von Modellen** oder nach **Änderungsprozessen** gefragt wurden. Das unstrukturierte Antworten kann in der vorliegenden Studie nur zum Teil bestätigt werden. Hier zeigten einige Schülerinnen und Schüler im Gegensatz zur Studie von Grosslight et al. (1991) in ihren Antworten eine Schritt-für-Schritt-Prozedur. Dabei durchlaufen die Lernenden in beiden Teilkompetenzen sukzessive die Niveaus. Diese Prozedur stammt von den Lernenden selbst, weshalb davon auszugehen ist, dass sie den Lernenden plausibel und für sie nachvollziehbar ist und daher für eine Entwicklung von Modellkompetenz bedeutsam sein könnte. Aus diesem Grund beschreibt Beitrag 5 (Grünkorn & Hanauer, 2013b) ein Beispiel, wie diese Prozedur bei der Teilkompetenz **Testen von Modellen** im Unterricht umgesetzt werden kann.

7 Implikationen für die zukünftige Forschung

Im Rahmen dieses Projekts wurde unter anderem das Kompetenzmodell der Modellkompetenz inhaltlich geprüft (Grünkorn, Upmeyer zu Belzen et al., 2014), und dies stellt einen weiteren Baustein der empirischen Überprüfung des Kompetenzmodells dar. Andere Fragen, etwa die Dimensionalität oder die Graduierung des Kompetenzmodells (vgl. Krell & Krüger, 2011; Krell, 2013; Krell et al., 2012a; Terzer, 2013), wurden jedoch nicht abschließend geklärt. Zur Bearbeitung solcher Fragestellungen sind Studien mit einem probabilistischen Ansatz notwendig, für die der hier entwickelte Test noch nicht umfassend geeignet war. Hier gibt es grundsätzlich zwei Möglichkeiten: Zum einen die Kombination verschiedener Aufgabenformate in einer längsschnittlichen Erfassung von Modellkompetenz, wie sie von Patzke et al. (*in Druck*) mit den hier konstruierten Aufgaben im offenen Antwortformat zusammen mit MC-Items (Terzer, 2013) durchgeführt wird. Auch andere Studien, wie die von Schwarz und White (2005), Treagust et al. (2004) sowie van Driel und Verloop (1999), kombinierten unterschiedliche Aufgabenformate und nutzten die Synergieeffekte. Zum anderen wäre auch an eine Erweiterung des Aufgabenpools zu denken. Mit der Konstruktion weiterer Aufgaben könnten das Fähigkeitsspektrum besser abgedeckt und damit die Validität und Reliabilität des hier entwickelten Tests verbessert werden (vgl. Neumann, 2014; Rost, 2004). Mit probabilistischen Studien kann außerdem die Frage nach der Bedeutung der didaktisch orientierten Kategorien *Modell als Mittel der Zugänglichkeit* und *Modell als Mittel der Verständlichkeit/Kommunizierbarkeit* für die Entwicklung von Modellkompetenz geklärt werden. Ausgehend von den Befunden lässt sich vermuten, dass zwischen den fünf Teilkompetenzen der Modellkompetenz und den beiden didaktisch orientierten Perspektiven ein Zusammenhang besteht (Abb. 14). Zudem kann angenommen werden, dass die Kategorie *Modell als Mittel der Zugänglichkeit* mit der Kategorie *Modell als Mittel der Verständlichkeit/Kommunizierbarkeit* in Zusammenhang steht (Abb. 14). Der Grund hierfür ist darin zu suchen, dass die Zugänglichmachung die Voraussetzung dafür schafft, Phänomene zu verstehen und zu begreifen (Ingham & Gilbert, 1991). In Abbildung 14 ist dieses komplexe Strukturmodell skizziert. Die genannte Hypothese sowie die Wirkung einer intensiven Reflexion beider Perspektiven im Unterricht sollten im Rahmen einer Interventionsstudie überprüft werden.

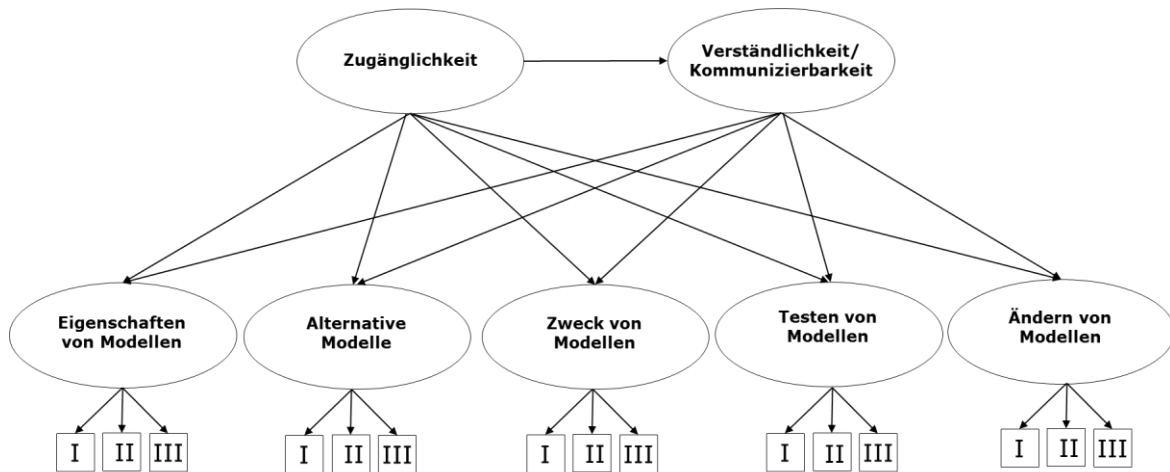


Abb. 14: Strukturmodell zum möglichen Zusammenhang der didaktisch orientierten Perspektiven mit den Teilkompetenzen¹²

Eine weitere offene Frage betrifft die Rolle des Parallelisierens von Modell und Original für die Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* bzw. für alle Teilkompetenzen des Kompetenzmodells. Bisherige Erkenntnisse legen nahe, dass das Parallelisieren eine Art Grundfähigkeit ist, welche die Performanz entweder bei den Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* oder bei allen Teilkompetenzen beeinflusst (Abb. 15). Um dies zu berücksichtigen, müssten die Beschreibungen des Niveaus II der Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* gegenüber der ursprünglichen Fassung verändert werden. Das Niveau II der Teilkompetenz *Testen von Modellen* wäre unter Hinzunahme des Kategoriensystems (Grünkorn, Upmeier zu Belzen et al., 2014) so anzupassen, dass diese Niveaustufe lediglich das Vergleichen zusammen mit der Beurteilung der Passung (Kategorie *Vergleich und Passung von Original und Modell*) umfasst. Das Niveau II der Teilkompetenz *Ändern von Modellen* könnte die Kategorien *Ändern bei neuen Erkenntnissen über das Original* und *Ändern bei Veränderungen des Originals* umfassen.

¹² Die römischen Ziffern stehen für die jeweiligen Niveaus der Teilkompetenz.

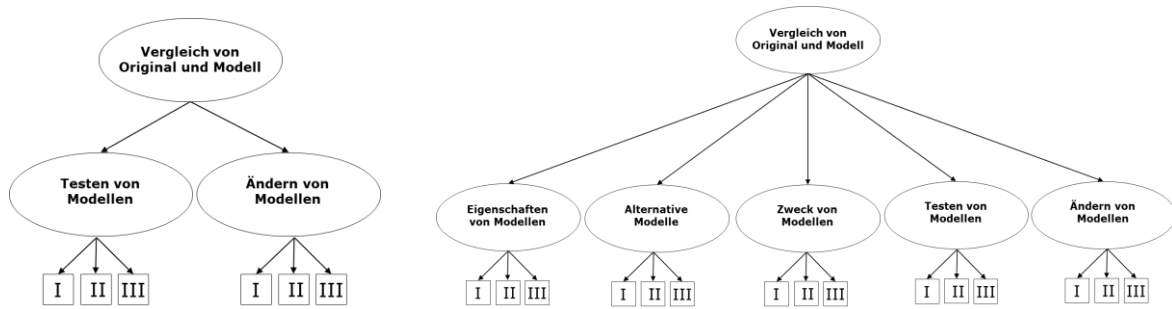


Abb. 15: Strukturmodelle zum möglichen Zusammenhang des Parallelisierens mit den Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* (links) bzw. allen Teilkompetenzen (rechts)¹³

Darüber hinaus sollte der entwickelte Leistungstest für den Einsatz in weiteren Schulformen als dem Gymnasium geprüft und gegebenenfalls optimiert werden, um Aussagen über ein breiteres Leistungsspektrum treffen zu können. Für den Einsatz in integrierten Sekundarschulen (Haupt-, Real- und Gesamtschulen) ist die Anwendbarkeit der entwickelten Instrumente nicht gesichert. Für Hauptschulen ist im Gegensatz zum Gymnasium eine weniger elaborierte Performanz der Schülerinnen und Schüler anzunehmen. Aus diesem Grund wäre es ratsam, die vorhandenen Aufgaben sprachlich zu vereinfachen und zusätzliche Aufgaben zu entwickeln, die auch von Personen mit geringeren Fähigkeiten erfolgreich bearbeitet werden können.

Außerdem ist es notwendig, einen Test zur Individualdiagnostik zu entwickeln, um adäquate Fördermaßnahmen für die Lernenden zu entwickeln und ihnen differenzierte Rückmeldungen zu geben. Ein solcher Test würde sich unter anderem durch eine Vielzahl von Aufgaben in variierenden Situationen auszeichnen, wodurch sich die Kompetenzen von Personen valide und sensitiv erfassen ließen (Klieme & Hartig, 2007). Sensitive Erfassung bedeutet, dass kleinschrittige Veränderungen erfasst und differenzierte Diagnosen ermöglicht werden. Erste Entwicklungsschritte sind bereits von Studien für die Teilkompetenzen *Eigenschaften von Modellen* (Barnetz, 2011), *Alternative Modelle* (Hanauer, 2012), *Testen von Modellen* (Pitthan, 2012) und *Ändern von Modellen* (Matheisius, 2012) umgesetzt worden. Sie entwickelten auf der Basis des Kategoriensystems zu Modellen und zur Modellbildung (Grünkorn, Upmeyer zu Belzen et al., 2014) Aufgaben im geschlossenen Format und berücksichtigen damit die verschiedenen Perspektiven von Lernenden. Gegenwärtig arbeiten Gogolin und

¹³ Die römischen Ziffern stehen für die jeweiligen Niveaus der Teilkompetenz.

Krüger (*in Druck*) an der Weiterentwicklung und Überprüfung des Tests zur Individualdiagnose.

Darüber hinaus wäre es möglich, die hier entwickelten Aufgaben im offenen Antwortformat in einem Test zur Individualdiagnostik einzusetzen, weil mit ihnen auch verschiedene Perspektiven auf Modelle und auf die Modellbildung erfasst werden können. Um den Bearbeitungsaufwand für eine Diagnose zu minimieren, könnte anstatt des herkömmlichen Papier-und-Bleistift-Tests ein computerbasiertes adaptives Testverfahren (CAT) genutzt werden (z. B. Frey, 2007). Bei einem solchen Verfahren kann beispielsweise die Aufgabenauswahl automatisch angepasst werden (adaptives Testen), wodurch sich die Testzeit reduziert und die Messeffizienz erhöht wird (z. B. Bjorner, Kosinski & Ware, 2003; Frey, 2007). Auch hier wäre eine Kombination mit MC-Items (Terzer, 2013), FC-Items (Krell & Krüger, 2010; Krell, Czeskleba et al., 2012) und den Items von Gogolin und Krüger (*in Druck*) denkbar. Für einen solchen Einsatz müssten alle Aufgaben kalibriert und in einem computerisierten Testformat (z. B. CBA ItemBuilder) administriert werden, und es müsste überprüft werden, ob der computerisierte Itempool für eine zuverlässige und valide Kompetenzmessung geeignet ist (Bjorner et al., 2003; van der Linden, 2005).

Literaturverzeichnis

- Abbot, A. (2005). Cell biology: Hopping fences. *Nature*, **433**, 680-683.
- Abel, G. (2008). Modell und Wirklichkeit. In U. Dirks & E. Knobloch (Hrsg.), *Modelle* (S. 31–46). Frankfurt a. M.: Peter Lang.
- Adams, R. J. & Wu, M. L. (2002). *PISA 2000 Technical Report*. Paris: Organisation for Economic Co-operation and Development (OECD).
- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention* (3. korrigierte, aktualisierte und überarbeitete Aufl.). Berlin: Springer.
- Bailer-Jones, D. M. (2002). Scientists' thoughts on scientific models [Electronic Version]. *Perspectives on Science*, **10** (3), 275-301.
- Barnetz, N. (2011). *Erfassung von Modellkompetenz bei Schülern mit Aufgaben im Multiple-Choice-Format in der Teilkompetenz „Eigenschaften von Modellen“*. Unveröffentlichte Bachelorarbeit, Freie Universität Berlin.
- Baumert, J., Stanat, P. & Demmrich A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15-68). Opladen: Leske & Budrich.
- Bayrhuber, H., Bögeholz, S., Eggert, S., Elster, D., Grube, C. & Höble, C. et al. (2007). Biologie im Kontext - Erste Forschungsergebnisse. *Der mathematische und naturwissenschaftliche Unterricht*, **60** (5), 304-312.
- Bayrhuber, H., Bögeholz, S., Elster, D., Hammann, M., Hössle, C., Lücken, M., Mayer, J., et al. (2007). Biologie im Kontext - Ein Programm zur Kompetenzförderung durch Kontextorientierung im Biologieunterricht und zur Unterstützung von Lehrerprofessionalisierung. *Der mathematische und naturwissenschaftliche Unterricht*, **60** (5), 282-286.
- Beerenwinkel, A. & Parchmann, I. (2008). Metadiskussion über Modelle. Historische Aspekte als Impuls. *Praxis der Naturwissenschaften – Chemie in der Schule*, **4** (57), 13–16.
- Bennett, R. E. & Ward, W. C. (Hrsg.). (1993). *Construction versus choice in cognitive measurement. Issues in constructed response, performance testing, and portfolio assessment*, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bjorner, J. B., Kosinski, M. & Ware, J. E. Jr. (2003). Calibration of an item pool for assessing the burden of headaches: An application of item response the-

- ory to the Headache Impact Test (HIT). *Quality of Life Research*, **12** (8), 913-933.
- Black, P. & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, **21** (1), 5-31.
- Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application an EM algorithm. *Psychometrika*, **46** (4), 443-459.
- Boersma, K. T. & Geraedts, C .L (2012). The interpretation of students' Lamarckian explanations [Abstract]. *Conference Programme Booklet of the 9th ERIDOB Conference*. (S. 54). Berlin: European Researchers in Didactics of Biology, 9th ERIDOB Conference.
- Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler (4., überarb. Aufl.)*. Berlin: Springer.
- Boulter, C. J. & Buckley, B. C. (2000). Constructing a typology of models for science education. In J. K. Gilbert & C. J. Boulter (Hrsg.), *Developing models in science education* (S. 41-57). Dordrecht: Kluwer Academic.
- Bühner, M. (2010). *Einführung in die Test- und Fragebogenkonstruktion (3., aktualisierte Aufl.)*. München: Pearson Studium.
- Bybee, R. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (Hrsg.), *Scientific Literacy* (S. 37-68). Kiel: IPN.
- Carey, S., Evans, R., Honda, M., Jay, E. & Unger, C. (1989). "An experiment is when you try it and see if it works": a study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, **11** (5), 514-529.
- Carstensen, C. H., Frey, A., Walter, O. & Knoll, S. (2007). Technische Grundlagen des dritten internationalen Vergleichs. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 367-390). Münster: Waxmann.
- Ciechorski, P. (2009). *Modellkompetenz im Teilbereich „Eigenschaften von Modellen“*. Unveröffentlichte Masterarbeit, Freie Universität Berlin.

- Cizek, G. J. & Bunch, M. B. (2007). *Standard-setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22 (9), 1041-1053.
- Colbourn, C. J. & Dinitz, J. H. (1996). *The CRC Handbook of Combinatorial Designs*. Boca Raton, FL: CRC Press.
- Coll, R. K., France, B. & Taylor, I. (2005). The role of models/and analogies in science education: implications from research. *International Journal of Science Education*, 27 (2), 183-198.
- Connell, M. W., Sheridan, K. & Gardner, H. (2003). On abilities and domains. In R. J. Sternberg & E. L. Grigorenko (Hrsg.), *The Psychology of Abilities, Competencies, and Expertise* (S. 126–155). Cambridge, UK: Cambridge University Press.
- Crawford, B. A. & Cullin, M. (2005). Dynamic assessments of preservice teachers' knowledge of models and modelling. In K. Boersma, M. Goedhart, O. de Jong & H. Eijkelhof (Hrsg.), *Research and the Quality of Science Education* (S. 309-323). Dordrecht: Springer.
- Crawford, B. A. & Cullin, M. J. (2004). Supporting prospective teachers' conceptions of modelling in science. *International Journal of Science Education*, 26 (11), 1379-1401.
- De Jong, O. & van Driel, J. H. (2001). *Developing preservice teachers' content knowledge and PCK of models and modelling*. Paper presented at the National Association for Research in Science Teaching Annual Meeting, St. Louis, MO, 25-28 March.
- Die Welt (2013, 7. Januar). *Cameron will Kompetenzen von EU zurückholen*. *Die Welt*, Zugriff am 07.01.2013. Verfügbar unter <http://www.welt.de/politik/ausland/article112436145/Cameron-will-Kompetenzen-von-EU-zurueckholen.html>
- Eggert, S. & Bögeholz, S. (2010). Students' use of decision-making strategies with regard to socioscientific issues - An application of the Rasch partial credit model. *Science Education*, 94 (2), 230-258.
- Eggert, S. (2008). *Bewertungskompetenz für den Biologieunterricht - Vom Modell zur empirischen Überprüfung*. Dissertation. Zugriff am 01.10.2012. Verfügbar unter <http://webdoc.sub.gwdg.de/diss/2008/eggert/eggert.pdf>

- Eggert, S., Bögeholz, S., Watermann, R. & Hasselhorn, M. (2010). Förderung von Bewertungskompetenz im Biologieunterricht durch zusätzliche metakognitive Strukturierungshilfen beim Kooperativen Lernen – Ein Beispiel für Veränderungsmessung. *Zeitschrift für Didaktik der Naturwissenschaften*, **16**, 299–314.
- Einhaus, E. (2007). *Schülerkompetenzen im Bereich Wärmelehre - Entwicklung eines Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells fachbezogener Kompetenzen*. Berlin: Logos.
- Fleige J., Seegers, A., Upmeier zu Belzen, A. & Krüger, D. (Hrsg.) (2012a). Modellkompetenz im Biologieunterricht 7-10. Donauwörth: Auer.
- Fleige, J., Seegers, A., Upmeier zu Belzen, A. & Krüger, D. (2012b). Förderung von Modellkompetenz im Biologieunterricht. *Der mathematische und naturwissenschaftliche Unterricht*, **65 (1)**, 19–28.
- Fleischer, J., Koeppen, K., Kenk, M., Klieme E. & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms [Sonderheft]. *Zeitschrift für Erziehungswissenschaft*, **16**, 5-22.
- Flick, U. (2002). *Qualitative Sozialforschung – Eine Einführung*. Reinbek: Rowohlt.
- Frey, A. (2007). Adaptives Testen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 261-278). Berlin, Heidelberg: Springer.
- Frey, A. (2014). Validität: Internationaler Forschungsstand und Umsetzung in Deutschland [Abstract]. 2. *Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF)*, 367. Zugriff am 28.02.2014. Verfügbar unter <http://www.gebf2014.de/abstractband>
- Frey, A. J., Hartig, J. & Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and Practice. *Educational Measurement*, **28 (3)**, 39–53.
- Frigg, R. & Hartmann, S. (2012). Models in science. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy*. Zugriff am 01.10.2012. Verfügbar unter <http://plato.stanford.edu/archives/fall2012/entries/models-science/>
- Füssenich, I., Granzer, D., Grassmann, M., Pech, D., Sandfuchs, U., Schilcher, A. et al. (Hrsg.) (2013). Modellhaft. Aufbau von Modellkompetenzen im Sa-

- chunterricht. *Grundschule*, **45** (6). Braunschweig: Bildungshaus Schulbuchverlage.
- Giere, R. N. (1988). *Explaining science: A cognitive approach. Science and its conceptual foundations*. Chicago, IL: University of Chicago Press.
- Giere, R. N. (2001). A new framework for teaching scientific reasoning. *Argumentation*, **15** (1), 21–33.
- Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, **71** (5), 742-752.
- Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, **172** (2), 269–281.
- Giere, R. N., Bickle, J. & Mauldin, R. F. (2006). *Understanding scientific reasoning*. Belmont, CA: Thomson Wadsworth.
- Giesbrecht, F. G. & Gumpertz, M. L. (2004). *Planning, Construction, and Statistical Analysis of Comparative Experiments*. Hoboken, NJ: Wiley.
- Gilbert, J. K. & Boulter, C. J. (Hrsg.). (2000). *Developing Models in Science Education*. Dordrecht: Kluwer.
- Gilbert, S. W. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, **28** (1), 73-79.
- Gläser-Zikuda, M. (2005). Qualitative Inhaltsanalyse in der Lernstrategie- und Lernemotionsforschung. In P. Mayring & M. Gläser-Zikuda (Hrsg.), *Die Praxis der Qualitativen Inhaltsanalyse* (S.64-85). Weinheim: Beltz UTB.
- Gogolin, S. & Krüger, D. (in Druck). Nature of models - Entwicklung von Diagnoseaufgaben. In *Lehr- und Lernforschung in der Biologiedidaktik* (Bd. 6).
- Greca, I. M. & Moreira, M. A. (2000). Mental models, conceptual models and modelling. *International Journal of Science Education*, **22** (1), 1-11.
- Gropengießer, H. (2003). *Wie man Vorstellungen der Lerner verstehen kann. Lebenswelten, Denkwelten, Sprechwelten*. Beiträge zur Didaktischen Rekonstruktion (Bd. 4.) Oldenburg: Didaktisches Zentrum.
- Gropengießer, H. (2005). Qualitative Inhaltsanalyse in der fachdidaktischen Lehr-Lern-Forschung. In P. Mayring & M. Gläser-Zikuda (Hrsg.), *Die Praxis der Qualitativen Inhaltsanalyse* (S. 172-189). Weinheim: Beltz UTB.
- Grosslight, L., Unger, C., Jay, E. & Smith, C. L. (1991). Understanding models and their use in science: conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, **28** (9), 799-822.

- Grube, C. (2011). *Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung: Untersuchung der Struktur und Entwicklung des wissenschaftlichen Denkens bei Schülerinnen und Schülern der Sekundarstufe I*. Dissertation. Zugriff am 01.10.2012. Verfügbar unter <http://nbn-resolving.de/urn:nbn:de:hebis:34-2011041537247>
- Grünkorn, J. & Fleige, J. (2012). Bau und Funktion der Fischhaut. In J. Fleige, A. Seegers, A. Upmeyer zu Belzen & D. Krüger (Hrsg.), *Modellkompetenz im Biologieunterricht Klasse 7-10. Phänomene begreifbar machen – in 11 komplett ausgearbeiteten Unterrichtseinheiten* (S. 23-28). Donauwörth: Auer.
- Grünkorn, J. & Hanauer, N. (2013a). Fischhaut [Materialien]. *Grundschule*, **45**(6), Beilage, XIII-XVI.
- Grünkorn, J. & Hanauer, N. (2013b). Schwimmen wie Fische im Wasser – Untersuchung der Schleimschicht von Fischen. *Grundschule*, **45** (6), 28-29.
- Grünkorn, J. & Krüger, D. (2012). Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells zur Modellkompetenz. In U. Harms & F. X. Bogner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (Bd. 5, S. 9-27). Innsbruck: Studienverlag.
- Grünkorn, J., Hänsch, J., Upmeyer zu Belzen, A. & Krüger, D. (2012). Determination of students' model competence using open-ended and hands-on tasks. In C. Bruguière, A. Tiberghien & P. Clément (Eds.), *E-Book Proceedings of the ESERA 2011 Conference: Science learning and Citizenship. Part 5* (co-ed Maurines, L., & Redfors, A.), (S. 39-45) Lyon, France: European Science Education Research Association.
- Grünkorn, J., Lotz, A. & Terzer, E. (2014). Erfassung von Modellkompetenz im Biologieunterricht. *Der mathematische und naturwissenschaftliche Unterricht*, **67** (3), 132-138.
- Grünkorn, J., Upmeyer zu Belzen, A. & Krüger, D. (2011). Design and test of open-ended tasks to evaluate a theoretical structure of model competence. In A. Yarden & G. S. Carvalho (Hrsg.), *Authenticity in Biology Education. Benefits and Challenges* (S. 53-65). Braga, Portugal: CIEC Universidade do Minho.
- Grünkorn, J., Upmeyer zu Belzen, A. & Krüger, D. (2014). *Assessing students' understandings of biological models and their use in science to evaluate a*

- theoretical framework. *International Journal of Science Education*, 1-34. doi: 10.1080/09500693.2013.873155
- Günther, J. (2006). *Lehrerfortbildung über die Natur der Naturwissenschaften: Studien über das Wissenschaftsverständnis von Grundschullehrkräften*. Berlin: Logos.
- Günther, S. L., Fleige, J., Upmeyer zu Belzen, A. & Krüger, D. (2014, Februar). *Förderung professioneller Kompetenzen Lehramtsstudierender im Umgang mit Modellen im Biologieunterricht*. Poster auf der 16. Internationale Frühjahrsschule der Fachsektion Didaktik der Biologie im VBIO, 24.-27.02.2014, Trier.
- Halloun, I. (2001). *Student views about science: A comparative survey*. Zugriff am 12.01.2012. Verfügbar unter <http://modeling.asu.edu/R&E/IHalloun/VASS-2001Monograph.pdf>
- Hammann, M. & Jördens, J. (2014). Offene Aufgaben codieren. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 169-178). Berlin: Springer.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle. Merkmale und ihre Bedeutung – dargestellt anhand von Kompetenzen beim Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht*, 57 (4), 196-203.
- Hammann, M. (2006). Kompetenzförderung und Aufgabenentwicklung. *Der mathematische und naturwissenschaftliche Unterricht*, 59 (2), 85-95.
- Hanauer, N. (2012). *Entwicklung und Testung von geschlossenen Aufgaben zur Erfassung von Modellkompetenz für die Teilkompetenz „Alternative Modelle“*. Unveröffentlichte Masterarbeit, Freie Universität Berlin.
- Hänsch, J. & Upmeyer zu Belzen, A. (2012, März). *Hands-On Aufgaben zur Erfassung und Förderung von Modellkompetenz*. Vortrag auf der 14. Frühjahrsschule der Fachsektion Didaktik der Biologie, Bremen.
- Harrison, A. G. & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, 22 (9), 1011-1026.
- Harrison, A. G. (2001). How do teachers and textbook writers model scientific ideas for students? *Research in Science Education*, 31 (3), 401-435.
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127-143). Berlin: Springer.

- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung* (S. 83-99). Weinheim: Beltz.
- Hartig, J. (2008). Kompetenzen als Ergebnisse von Bildungsprozessen. In N. Jude, J. Hartig & E. Klieme (Hrsg.), *Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden* (S. 15-25). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Hartig, J. (2009, August). *IRT-Skalierung mit ConQuest*. Workshop auf der Internationalen Summer School der Deutschen Gesellschaft für Erziehungswissenschaft, Lingen.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143-171). Berlin: Springer.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77* (1), 81-112.
- Helmke, A. & Hosenfeld, I. (2004). Vergleichsarbeiten – Standards – Kompetenzstufen: Begriffliche Klärung und Perspektiven. In M. Wosnitza, A. Frey & R. S. Jäger (Hrsg.), *Lernprozess, Lernumgebung und Lerndiagnostik* (S. 56-75). Landau: Verlag Empirische Pädagogik.
- Helmke, A., Goellner, R., Kleinbub, I., Schrader, F.-W. & Wagner, W. (2008). Standards - Motor oder Bremse der Unterrichtsentwicklung? *Seminar - Lehrerbildung und Schule. Kompetenzerwerb in der Lehrerbildung, 14* (1), 44-61.
- Henze, I., van Driel, J. H. & Verloop, N. (2007). Science teachers' knowledge about teaching models and modelling in the context of a new syllabus on public understanding of science. *Research in Science Education, 37*, 99-122.
- Hodson, D. (1992). In search of a meaningful relationship: An exploration of some issues relating to integration in science and science education. *International Journal of Science Education, 14* (5), 541-562.
- Hodson, D. (1993). Re-thinking old ways: Towards a more critical approach to practical work in school science. *Studies in Science Education, 22* (1), 85-142.
- Hofer, B. K. & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67* (1), 88-140.

- Ingham, A. M. & Gilbert, J. K. (1991). The use of analogue models by students of chemistry at higher education level. *International Journal of Science Education*, **13** (2), 193-202.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Jonkisz, E., Moosbrugger, H. & Brandt, H. (2012). Planung und Entwicklung von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 28–74). Berlin: Springer.
- Justi, R. S. & Gilbert, J. K. (2001, March). *Teachers' views on models and modelling in science education*. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, St. Louis, MI.
- Justi, R. S. & Gilbert, J. K. (2002). **Modelling, teachers' views on the nature of modelling, and implications for the education of modellers.** *International Journal of Science Education*, **24** (4), 369–387.
- Justi, R. S. & Gilbert, J. K. (2003). **Teachers' views on the nature of models.** *International Journal of Science Education*, **25** (11), 1369-1386.
- Justi, R. S. & Gilbert, J. K. (2006). The role of analog models in the understanding of the nature of models in chemistry. In P. J. Aubusson, A. G. Harrison & S. M. Ritchie (Hrsg.), *Metaphor and Analogy in Science Education* (S. 119–130). Dordrecht: Springer.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, **38** (4), 319–342.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, **50** (1), 1-73.
- Kattmann, U. (2006). Modelle. In H. Gropengießer & U. Kattmann (Hrsg.), *Fachdidaktik Biologie: Die Biologiedidaktik* (7. Auflage) (S. 330–339). Köln: Aulis.
- Kauertz, A. (2014). Entwicklung eines Rasch-skalierten Leistungstests. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 341-353). Berlin: Springer.
- Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E. & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, **16**, 135-153.

- Kelava, A. & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 75–102). Berlin: Springer.
- Klauer, K. (1987). Kriteriumsorientiertes Testen. Der Schluss auf den Itempool. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *8*, 141–147.
- Klieme, E. & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs [Sonderheft]. *Zeitschrift für Erziehungswissenschaft*, *8*, 11–29.
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: Beschreibung eines neu eingerichteten Schwerpunktprogramms der Deutschen Forschungsgemeinschaft. *Zeitschrift für Pädagogik*, *52* (6), 876–903.
- Klieme, E. & Steinert, B. (2004). Einführung der KMK-Bildungsstandards: Zielsetzungen, Konzeptionen und Einführung in den Schulen am Beispiel der Mathematik. *Der mathematische und naturwissenschaftliche Unterricht*, *57* (3), 132–137.
- Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*, *56* (6), 10–13.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2007). *Zur Entwicklung nationaler Bildungsstandards – Expertise*. Bonn, Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Klieme, E., Baumert, J., Köller, O. & Bos, W. (2000). Mathematische und naturwissenschaftliche Grundbildung: Konzeptionelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III* (Band 1, S. 85–133). Opladen: Leske + Budrich.
- Klieme, E., Hartig, J. & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of Competencies in Educational Contexts* (S. 3–22). Cambridge, UK: Hogrefe & Huber.
- Klieme, E., Maag Merki, K. & Hartig, J. (2007). Kompetenzbegriff und Bedeutung von Kompetenzen im Bildungswesen. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiag-*

- nostik* (S. 5–15). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Köller, O. (2008). Bildungsstandards: Verfahren und Kriterien bei der Entwicklung von Messinstrumenten. *Zeitschrift für Pädagogik*, *54* (2), 163–173.
- Köller, O. (2009). Bildungsstandards. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 529-548). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Koponen, I. T. (2007). Models and modelling in physics education: A critical re-analysis of philosophical underpinnings and suggestions for revisions. *Science & Education*, *16*, 751–773.
- Koslowski, F. (2009). *Modellkompetenz im Teilbereich „Modellalternativen“*. Unveröffentlichte Masterarbeit, Freie Universität Berlin.
- Kramer, G. (2009). *Entwicklung und Überprüfung eines Strukturmodells der fachlichen Kommunikationskompetenz im Biologieunterricht*. Dissertation. Zugriff am 01.10.2012. Verfügbar unter http://eldiss.uni-kiel.de/macau/receive/dissertation_diss_00004782
- Krell, M. & Krüger, D. (2011). Forced Choice-Aufgaben zur Evaluation von Modellkompetenz im Biologieunterricht: Empirische Überprüfung konstrukt- und merkmalsbezogener Teilkompetenzen. *Erkenntnisweg Biologiedidaktik*, *10*, 53-68. Zugriff am 01.02.2013. Verfügbar unter <http://www.bcp.fu-berlin.de/biologie/arbeitsgruppen/didaktik/Erkenntnisweg/2011/index.html>
- Krell, M. (2012). Using polytomous IRT models to evaluate theoretical levels of understanding models and modeling in biology education. *Science Education Review Letters, Theoretical Letters 2012*, 1-5. Zugriff am 01.02.2013. Verfügbar unter http://edoc.hu-berlin.de/serl/2012-1/PDF/2012_1.pdf
- Krell, M. (2013). *Wie Schülerinnen und Schüler biologische Modelle verstehen: Erfassung und Beschreibung des Modellverstehens von Schülerinnen und Schülern der Sekundarstufe I*. Berlin: Logos.
- Krell, M., & Krüger, D. (2010). Diagnose von Modellkompetenz: Deduktive Konstruktion und Selektion von geschlossenen Items. *Erkenntnisweg Biologiedidaktik*, *9*, 23–37.
- Krell, M., Czeskleba, A. & Krüger, D. (2012). Validierung von Forced Choice-Aufgaben durch Lautes Denken. *Erkenntnisweg Biologiedidaktik*, *11*, 53-70.

- Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2014a). **Students' levels of understanding models and modelling in biology: Global or aspect-dependent?** *Research in Science Education*, **44**, 109-132.
- Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2014b). How year 7 to year 10 students categorise models: Moving towards a student-based typology of biological models. In D. Krüger & M. Ekborg (Hrsg.), *Research in biological education* (S. 117-131).
- Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2012a). **Assessment of students' concepts of models and modeling: Empirical evaluation of a model of model competence.** In C. Bruguière, A. Tiberghien, & P. Clément (Eds.), *E-book proceedings of the ESERA 2011 conference. Part 5: Nature of science, history, philosophy, sociology of science* (S. 68–74). Zugriff am 06.05.2014. Verfügbar unter <http://www.esera.org/media/ebook/ebook-esera2011.pdf>
- Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2012b). **Students' understanding of the purpose of models in different biological contexts.** *International Journal of Biology Education*, **2** (2), 1-34. Zugriff am 01.02.2013. Verfügbar unter http://www.ijobed.com/2_2/Moritz-2012.pdf
- Krell, M., Upmeier zu Belzen, A. & Krüger, D. (2014c). **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. Strand 6. Nature of science: History, philosophy and sociology of science.* Nicosia, Cyprus: European Science Education Research Association. Zugriff am 06.05.2014. Verfügbar unter http://www.esera.org/media/esera2013/Moritz_Krell_07Feb2014.pdf
- Krosnick, J. A. (1999). Survey research. *Annual review of Psychology*, **50**, 537-567.
- Krüger, D. & Riemeier, T. (2014). Die qualitative Inhaltsanalyse – eine Methode zur Auswertung von Interviews. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 133-145). Berlin: Springer.
- Krüger, D. (2007). Die Conceptual Change-Theorie. In D. Krüger & H. Vogt. (Hrsg.), *Handbuch der Theorien in der biologiedidaktischen Forschung* (S. 81-92). Berlin: Springer.

- Kunter, M. & Klusmann, U. (2010). Kompetenzmessung bei Lehrkräften – Methodische Herausforderungen. *Unterrichtswissenschaft*, **38** (1), 68-86.
- Kusumi, A. & Sako, Y. (1996). Cell surface organization by the membrane skeleton. *Current Opinion in Cell Biology*. **8** (4), 566–574.
- Lachmayer, S. (2008). *Entwicklung und Überprüfung eines Strukturmodells der Diagrammkompetenz für den Biologieunterricht*. Dissertation. Zugriff am 01.10.2012. Verfügbar unter http://eldiss.uni-kiel.de/macau/receive/dissertation_diss_00003041
- Laux, A. (2013, 9. Januar). Im Aufsichtsrat fehlt die sportliche Kompetenz. *Die Welt*, Zugriff am 09.01.2013. Verfügbar unter <http://www.welt.de/regionales/hamburg/article112640619/Im-Aufsichtsrat-fehlt-die-sportliche-Kompetenz.html>
- Leisner, A. & Mikelskis, H. F. (2004). Erwerb metakonzeptueller Kompetenz durch ein systematisches Lernen über Modelle. In A. Pitton (Hrsg.), *Chemie- und physikdidaktische Forschung und naturwissenschaftliche Bildung* (S. 120–122). Münster: LiT.
- Leisner-Bodenthin, A. (2006). Zur Entwicklung von Modellkompetenz im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, **12**, 91–108.
- Leuders, T. (2014). Modellierungen mathematischer Kompetenzen – Kriterien für eine Validitätsprüfung aus fachdidaktischer Sicht. *Journal für Mathematik-Didaktik*, **35** (1), 7-48.
- Leutner, D., Fleischer, J., Spoden, C. & Wirth, J. (2007). Landesweite Lernstandserhebungen zwischen Bildungsmonitoring und Individualdiagnostik [Sonderheft]. *Zeitschrift für Erziehungswissenschaften*, **8**, 149–167.
- Lienert, G. A. & Ratz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Louca, L. T., Zacharia, Z. C., Michael, M. & Constantinou, C. P. (2011). Objects, entities, behaviors and interactions: A typology of student-constructed computer-based models of physical phenomena. *Journal of Educational Computing Research*, **44** (2), 173-202.
- Mahr, B. (2003). Modellieren: Beobachtungen und Gedanken zur Geschichte des Modellbegriffs. In S. Krämer & H. Bredekamp (Hrsg.), *Bild, Schrift, Zahl* (S. 59–86). München: Fink.

- Mahr, B. (2004). *Das Wissen im Modell* (KIT-Report Nr. 150). TU-Berlin. Zugriff am 01.10.2012. Verfügbar unter <https://www.flp.tu-berlin.de/fileadmin/fg53/KIT-Reports/r150.pdf>
- Mahr, B. (2008). Ein Modell des Modellseins: Ein Beitrag zur Aufklärung des Modellbegriffs. In U. Dirks & E. Knobloch (Hrsg.), *Modelle* (S. 187–218). Frankfurt a. M.: Peter Lang.
- Mahr, B. (2009). Information science and the logic of models. *Software & Systems Modeling*, **8** (3), 365-383.
- Marquard, O. (1981). *Abschied vom Prinzipiellen. Philosophische Studien*. Stuttgart: Reclam.
- Martinez, M. E. (1991). A comparison of multiple-choice and constructed figural response items. *Journal of Educational Measurement*, **28** (2), 131-145.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, **47**, 149-174.
- Mathesius, S. (2012). *Entwicklung und Testung von geschlossenen Aufgaben zur Erfassung von Modellkompetenz im Biologieunterricht für die Teilkompetenz „Ändern von Modellen“*. Unveröffentlichte Masterarbeit, Freie Universität Berlin.
- MAXQDA (Version 10) [Computer Software]. Berlin: VERBI Software. Zugriff am 09.04.2010. Verfügbar unter <http://www.maxqda.de/>
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 177–186). Berlin: Springer.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (11., aktualisierte und überarb. Aufl.). Weinheim: Beltz.
- McClelland, D. C. (1973). Testing for Competence Rather Than for "Intelligence". *American Psychologist*, **28** (1), 1–14.
- Meisert, A. (2008). Vom Modellwissen zum Modellverständnis: Elemente einer umfassenden Modellkompetenz und deren Fundierung durch lernerseitige Kriterien zur Klassifikation von Modellen. *Zeitschrift für Didaktik der Naturwissenschaften*, **14**, 243–261.
- Meisert, A. (2009). Modelle in der Biologie: Wie lässt sich im Unterricht ein Verständnis für ihre Bedeutung fördern? Der *mathematische und naturwissenschaftliche Unterricht*, **62** (7), 424–430.

- Meyer, H. (2012, Februar). *Kompetenzorientierung allein macht noch keinen guten Unterricht*. Vortrag auf der didacta 2012, Hannover.
- Mikelskis-Seifert, S. & Fischler, H. (2003). Die Bedeutung des Denkens in Modellen bei der Entwicklung von Teilchenvorstellungen – Stand der Forschung und Entwurf einer Unterrichtskonzeption. *Zeitschrift für die Didaktik der Naturwissenschaften*, *9*, 75–88.
- Mikelskis-Seifert, S. & Kasper, L. (2011). Modellieren in der Physik, im Alltag und im Unterricht. Hintergründe und unterrichtliche Orientierung zum Thema Modelle [Themenheft „Modelle“]. *Unterricht Physik*, *122*, 4-12.
- Mittelstraß, J. (2004). *Enzyklopädie Philosophie und Wissenschaftstheorie*. Stuttgart: Metzler.
- Moosbrugger, H. (2012). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 103–117). Berlin: Springer.
- Morrison, M. & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Hrsg.), *Models as mediators. Perspectives on natural and social sciences* (S. 10–37). Cambridge, UK: Cambridge University Press.
- National Research Council (2012). *A framework for K-12 science education: practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Hrsg.), *Cognitive models of science* (S. 3–45). Minneapolis, MN: University of Minnesota Press.
- Neuhaus, B. & Braun, E. (2007). Testkonstruktion und Testanalyse – praktische Tipps für empirisch arbeitende Didaktiker und Schulpraktiker. In H. Bayrhuber, D. Elster, D. Krüger & H. J. Vollmer (Hrsg.), *Forschungen zur Fachdidaktik: Kompetenzentwicklung und Assessment* (S. 135–164). Innsbruck: Studienverlag.
- Neumann, K. (2014). Rasch-Analyse naturwissenschaftsbezogener Leistungstests. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 355-369). Berlin: Springer.
- Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Hrsg.), *Mental models* (S. 7–14). Hillsdale, NJ: Lawrence Erlbaum.

- Odenbaugh, J. (2005). Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology. *Biology & Philosophy*, *20* (2-3), 231-255.
- Oh, P. S. & Oh, S. J. (2011). What teachers of science need to know about models: An overview. *International Journal of Science Education*, *33* (8), 1109–1130.
- Orsenne, J. & Upmeyer zu Belzen, A. (2012). Hands-On Aufgaben zur Erfassung und Förderung von Modellkompetenz im Biologieunterricht. In U. Harms & F. X. Bogner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (S. 29-44). Innsbruck: Studienverlag.
- Pant, H. A. (2013). Wer hat ein Nutzen von Kompetenzmodellen? [Sonderheft]. *Zeitschrift für Erziehungswissenschaft*, *16* (1), 71-79.
- Pant, H. A., Böhme, K. & Köller, O. (2012). Das Kompetenzkonzept der Bildungsstandards und die Entwicklung von Kompetenzstufenmodellen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik* (S. 49–55). Münster: Waxmann.
- Patzke, C. & Upmeyer zu Belzen, A. (2012). Entwicklung von Modellkompetenz und entsprechende Lerngelegenheiten im Biologieunterricht. In D. Elster, A. Schultz-Siatkowski & F. Wischmann (Hrsg.), *Tagungsband der 14. Frühjahrsschule 2012 der Fachdidaktik für Biologie, Biowissenschaften und Biomedizin* (S. 106-107). Bremen: Universität Bremen.
- Patzke, C., Krüger, D. & Upmeyer zu Belzen, A. (in Druck). Entwicklung von Modellkompetenz im Längsschnitt. In *Lehr- und Lernforschung in der Biologiedidaktik* (Bd. 6).
- Pellegrino, J. W., Chudowsky, N. & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: National Academic Press.
- Pitthan, T. (2012). *Entwicklung und Überprüfung von geschlossenen Aufgaben zur Erfassung von Modellkompetenz bei Schüler/-innen in der Teilkompetenz „Testen von Modellen“*. Unveröffentlichte Bachelorarbeit, Freie Universität Berlin.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. & Pekrun, R. (Hrsg.). (2008). PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich. Münster: Waxmann.

- Prenzel, M., Drechsel, B., Carstensen, C. H. & Ramm, G. (2004). PISA 2003 – eine Einführung. In M. Prenzel, J. Baumert, W. Blum, R. H. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 13-46). Münster: Waxmann.
- Prenzel, M., Häußler, P., Rost, J. & Senkbeil, M. (2002). Der PISA-Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten vorhersagen? *Unterrichtswissenschaft*, **30** (2), 120–135.
- Prenzel, M., Walter, O. & Frey, A. (2007). PISA misst Kompetenzen. Eine Republik auf Rindermann (2006): Was messen internationale Schulleistungsstudien? *Psychologische Rundschau*, **58** (2), 128-136.
- Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *The Danish Yearbook of Philosophy*, **14**, 58–93.
- Rauch, D. & Hartig, J. (2012). Interpretation von Testwerten in der IRT. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 253–264). Berlin: Springer.
- Reitschert, K. & Hößle, C. (2007). Wie Schüler ethisch bewerten: Eine qualitative Untersuchung zur Strukturierung und Ausdifferenzierung von Bewertungskompetenz in bioethischen Sachverhalten bei Schülern der Sek. I. *Zeitschrift für Didaktik der Naturwissenschaften*, **13**, 125–143.
- Robitzsch, A. (2013). Wie robust sind Struktur- und Niveaumodelle? Wie zeitlich stabil und über Situationen hinweg konstant sind Kompetenzen? [Sonderheft]. *Zeitschrift für Erziehungswissenschaft*, **16** (1), 41-45.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- Rotbain, Y., Marbach-Ad, G. & Stavy, R. (2006). Effect of bead and illustration models on high school students' achievement in molecular genetics. *Journal of Research in Science Teaching*, **43** (5), 500–529.
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, **12**, 45–66.
- Schmiemann, P. & Lücken, M. (2014). Validität – Misst mein Test, was er soll? In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 107-118). Berlin: Springer.
- Schmiemann, P. (2010). *Modellierung von Schülerkompetenzen im Bereich des biologischen Fachwissens*. Berlin: Logos.

- Schott, F. & Azizi Ghanbari, S. (2008). *Kompetenzdiagnostik, Kompetenzmodelle, kompetenzorientierter Unterricht. Zur Theorie und Praxis überprüfbarer Bildungsstandards. ComTrans – ein theoriegeleiteter Ansatz zum Kompetenztransfer als Diskussionsvorlage*. Münster: Waxmann.
- Schröder, C. & Wirth, I. (2012). *99 Tipps. Kompetenzorientiert unterrichten*. Berlin: Cornelsen.
- Schwarz, C. V. & White, B. Y. (2005). Metamodeling Knowledge: Developing Students' Understanding of Scientific Modeling. *Cognition and Instruction*, **23** (2), 165–205.
- 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. C. Alonzo, & A. W. Gotwals (Hrsg.), *Learning progressions in science. Current challenges and future directions* (S. 101–137). Rotterdam: Sense.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Acher, A., Fortus, D. et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners [Special Issue]. *Journal of Research in Science Teaching*, **46** (6), 632–654.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (Hrsg.) (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München Neuwied: Wolters Kluwer.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (Hrsg.) (2010). *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Köln: Wolters Kluwer.
- Senkbeil, M., Rost, J., Carstensen, C. H. & Walter, O. (2005). Der nationale Naturwissenschaftstest PISA 2003: Entwicklung und empirische Überprüfung eines zweidimensionalen Facettendesigns. *Empirische Pädagogik*, **19** (2), 166-189.
- Shavelson, R. J., Baxter, G. P. & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, **30** (3), 215–232.
- Simons, K. & Ikonen, E. (1997). Functional rafts in cell membranes. *Nature*, **387**, 569-572

- Simonton, D. K. (2003). Expertise, competence, and creative ability. The perplexing complexities. In R. J. Sternberg & E. L. Grigorenko (Hrsg.), *The Psychology of Abilities, Competencies, and Expertise* (S. 213–239). Cambridge, NY: Cambridge University Press.
- Singer, S. J. & Nicolson, G. L. (1972). The fluid mosaic model of the structure of cell membranes. *Science*, **175**, 720–731.
- Sins, P. H. M., Savelsbergh, E. R., van Joolingen, W. R. & van Hout-Wolters, B. H. A. M. (2009). The relation between students' epistemological understanding of computer models and their cognitive processing on a modelling task. *International Journal of Science Education*, **31** (9), 1205–1229.
- Smit, J. J. A. & Finegold, M. (1995). Models in physics: Perceptions held by final-year preservice physical science teachers studying at South African Universities. *International Journal of Science Education*, **17** (5), 621–634.
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Wien: Springer.
- Strike, K. A. & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Hrsg.), *Philosophy of science, cognitive psychology, and educational theory and practice* (S. 147–176). Albany, NY: SUNY Press.
- Suzuki, K., Ritchie, K., Kajikawa, E., Fujiwara, T. & Kusumi, A. (2005). Rapid hop diffusion of a G-protein-coupled receptor in the plasma membrane as revealed by single-molecule techniques. *Biophysical Journal*, **88** (5), 3659–3680.
- Terzer, E. (2013). *Modellkompetenz im Kontext Biologieunterricht – Empirische Beschreibung von Modellkompetenz mithilfe von Multiple-Choice Items*. Dissertation. Zugriff am 01.05.2013. Verfügbar unter <http://edoc.hu-berlin.de/dissertationen/terzer-eva-2012-12-19/PDF/terzer.pdf>
- Terzer, E., Hartig, J. & Upmeyer zu Belzen, A. (2013). Systematische Konstruktion eines Tests zu Modellkompetenz im Biologieunterricht unter Berücksichtigung von Gütekriterien. *Zeitschrift für Didaktik der Naturwissenschaften*, **19**, 51–76.
- Terzer, E., Patzke, C. & Upmeyer zu Belzen, A. (2012). Validierung von Multiple-Choice Items zur Modellkompetenz durch lautes Denken. In U. Harms & F. X. Bogner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (S. 45–62). Innsbruck: Studienverlag.

- Traub, R. E. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. In R. E. Bennett & W. C. Ward (Hrsg.). *Construction versus choice in cognitive measurement. Issues in constructed response, performance testing, and portfolio assessment* (S. 29-44). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Treagust, D. F., Chittleborough, G. D. & Mamiala, T. L. (2001). *Learning introductory organic chemistry: secondary students' understanding of the role of models and the development of scientific ideas*. Paper presented at American Educational Research Association, Seattle, WA.
- Treagust, D. F., Chittleborough, G. D. & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, *24* (4), 357–368.
- Treagust, D. F., Chittleborough, G. D. & Mamiala, T. L. (2004). Students' understanding of the descriptive and predictive nature of teaching models in organic chemistry. *Research in Science Education*, *34*, 1–20.
- Trier (2013). "Modell sind künstlich ..." – Schülervorstellungen zu Modellen in der Grundschule. *Grundschule*, *45* (6), 12-14.
- Trier, U. & Upmeier zu Belzen, A. (2009). „Wissenschaftler nutzen Modelle, um etwas Neues zu entdecken, und in der Schule lernt man einfach nur, dass es so ist.“: Schülervorstellungen zu Modellen. *Erkenntnisweg Biologiedidaktik* *8*, 23–38.
- Upmeier zu Belzen, A. & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, *16*, 41–57.
- van der Linden, W. J. (2005). A Comparison of Item-Selection Methods for Adaptive Tests with Content Constraints. *Journal of Educational Measurement*, *42* (3), 283-302.
- van der Valk, T., van Driel, J. & de Vos, W. (2007). Common characteristics of models in present-day scientific practice. *Research in Science Education*, *37* (4), 469–488.
- van Driel, J. H. & Verloop, N. (1999). Teachers' knowledge of models and modelling in Science. *International Journal of Science Education*, *21* (11), 1141–1153.
- van Driel, J. H. & Verloop, N. (2002). Experienced teacher's knowledge of teaching and learning of models and modeling in science education. *International Journal of Science Education*, *24* (12), 1255–1272.

- van Fraassen, B. C. (1980). *The scientific image*. Oxford: Clarendon Press.
- Volodin, N. A. & Adams, R. J. (1995). *Identifying and estimating a D-dimensional item response model*. Paper presented at the International Objective Measurement Workshop. Berkeley, CA: University of California.
- Vosniadou, S. (2002). Mental models in conceptual development. In L. Magnani & N. J. Nersessian (Hrsg.), *Model-based reasoning. Science, technology, values* (S. 353–368). New York: Kluwer Academic.
- Weinert, F. E. (1999). *Konzepte der Kompetenz*. Paris: Organisation for Economic Co-operation and Development (OECD).
- Weinert, F. E. (2001). Concepts of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Seattle, WA: Hogrefe & Huber.
- Weitzel, H. & Gropengießer, H. (2009). Vorstellungsentwicklung zur stammesgeschichtlichen Anpassung: Wie man Lernhindernisse verstehen und förderliche Lernangebote machen kann. *Zeitschrift für Didaktik der Naturwissenschaften*, *15*, 287-305.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (2007). Australian Council for Educational Research (ACER) *ConQuest Version 2.0: Generalised item response modelling software*. Camberwell, VIC: ACER Press.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (2012). Australian Council for Educational Research (ACER) *ConQuest Version 3.0.1* [Computer Software]. Camberwell, VIC: ACER Press.
- Zeybak, Y. (2010). *Diagnose von Modellkompetenz im Teilbereich „Zweck von Modellen“*. Unveröffentlichte Bachelorarbeit, Freie Universität Berlin.
- Zöfel, P. (2003). *Statistik für Psychologen: im Klartext*. München: Pearson Studium.

Publikationen dieser Arbeit

Für die Beiträge 1 bis 3 sind alle Erhebungen und Analysen von der Erst-Autorin durchgeführt worden. Die Manuskripte sind von ihr konzeptuell strukturiert und jeweils eine vollständige Version in Absprache mit der Ko-Autorin bzw. dem Ko-Autor verfasst worden. Nach einer Revision auf der Basis von Anmerkungen der Ko-Autorin bzw. des Ko-Autors wurden die Manuskripte eingereicht und nach den Kommentaren der Gutachterinnen und der Gutachter überarbeitet und vom Verlag angenommen sowie veröffentlicht.

Beitrag 4 ist gemeinsam von der Erst- und Dritt-Autorin konzeptuell strukturiert worden. Der Abschnitt „Schülerperspektiven auf Modelle und Modellbildung im Kontext Biologieunterricht“ beschreibt Ergebnisse aus der Forschungsarbeit der Erst-Autorin und ist von ihr verfasst. **Die Abschnitte „Instrumente zur Erfassung von Modellkompetenz in der Schule“ sowie „Einsatz der Instrumente“ sind von der Dritt-Autorin erarbeitet.** Die präsentierten Instrumente wurden von der Erst- und Dritt-Autorin gemeinsam erarbeitet. Die Erprobung der Instrumente führte der Zwei-Autor durch. Das eingereichte Manuskript wurde nach den Anmerkungen des Verlags nur geringfügig geändert, angenommen und veröffentlicht.

Die Beiträge 5 und 6 beschreiben Unterrichtskonzepte, in welche Implikationen aus den Studien der Erst-Autorin aufgegriffen wurden. Die Ideen und Konzeptionen entwickelte die Erst-Autorin und verfasste eine vollständige Version mit Absprache der Ko-Autorinnen. Die Ko-Autorinnen überprüften die Ideen und Konzeptionen hinsichtlich ihrer Praktikabilität im Unterricht. Die produzierten Textteile wurden nach den Anmerkungen der Ko-Autorinnen geringfügig geändert, vom Verlag angenommen und veröffentlicht.

Die folgenden Beiträge sind nach deren Einbindung im Rahmen der vorliegenden Dissertation geordnet.

Beitrag 1

Grünkorn, J., Upmeyer zu Belzen, A. & Krüger, D. (2011). Design and test of open-ended tasks to evaluate a theoretical structure of model competence. In A. Yarden & G.S. Carvalho (Eds.), *Authenticity in Biology Education. Benefits and Challenges* (S. 53-65). Braga, Portugal: CIEC Universidade do Minho.

5

DESIGN AND TEST OF OPEN-ENDED TASKS TO EVALUATE A THEORETICAL STRUCTURE OF MODEL COMPETENCE

Juliane Grünkorn¹, Annette Upmeier zu Belzen² and Dirk Krüger¹
¹Freie Universität Berlin, ²Humboldt Universität zu Berlin, Biology
Education (Berlin – Germany)
juliane.gruenkorn@fu-berlin.de

Abstract

Studies have shown that most students' conceptions of models differ from scientific conceptions (e.g. Grosslight, Unger, Jay & Smith, 1991). To describe scientific and everyday conceptions, Upmeier zu Belzen and Krüger (2010) developed a theoretical structure entailing two dimensions: 'knowledge about models' and 'modeling' which are subdivided into different aspects. This theoretical structure needs to be evaluated empirically.

This research project aims to operationalize and evaluate the theoretical structure using open-ended tasks. Fifteen tasks were designed and tested for understandability and consistency with the theoretical structure (N=912). The data were analyzed using qualitative content analysis (Mayring, 2003).

The article focuses on the findings concerning the dimension 'knowledge about models'. The analysis of student answers showed that difficulties mainly occur with the instructions and certain task contexts. Most student answers were consistent with the theoretical structure. The percentage distributions of the answers within each aspect corresponded to findings of other studies (e.g. Grosslight et al., 1991). One additional facet 'the uniqueness of models' was identified in student answers and needs to be considered. Moreover, students had alternative understandings of the term 'model', ranging from not perceiving certain objects as models to accepting everything as a model.

Keywords: models; model competence; theoretical structure of model competence; operationalization of a competence model; open-ended tasks.

1. Introduction

Models are an important part of scientific thinking and working methods. Thinking in models enables people to communicate about scientific topics and to gain flexible and transferable knowledge (Clement, 2000). Various studies have revealed that students reflect little on their thinking in and handling of models, and are not aware of the role models play in an epistemological process (e.g. Grosslight et al., 1991; Trier & Upmeier zu Belzen, 2009). Similar findings are given by PISA 2000 (Artelt et al., 2001) and 2003 (Prenzel et al., 2004), showing that students have difficulties in tasks where the handling of and thinking in models

are essential to solving the problem. To relate scientific and everyday conceptions, Upmeier zu Belzen and Krüger (2010) designed a theoretical structure of model competence which needs to be evaluated empirically. The aims of this research project are to operationalize and evaluate the theoretical structure using open-ended tasks.

This article focuses on the findings of the design and testing of open-ended tasks to evaluate the theoretical structure for the dimension 'knowledge about models'. In addition, results regarding the content evaluation of the theoretical structure are shown.

2. Theoretical background

2.1 Scientific, students' and teachers' conceptions of models

Many different objects, such as toy cars, computer simulations, and formulas, are seen as 'models'. Despite this broad concept of the term 'model', there is one thing all models have in common: they are carriers of ideas and conceptions (Mahr, 2009; Stachowiak, 1973). Gilbert (1991) defines scientific work as a construction of models which represent concepts of the world. Therefore, models are the product and method of science. In biology education, the modeled part of the world is often called the original (Kattmann, 2006).

The model is characterized by two different relationships to the original (Mahr, 2009): the creation of a model object by observing the original (model *of* something) and the application of a model object by testing it and drawing conclusions on the original (model *for* something) (Mahr, 2009). These characterizations, model object, model *of* something, and model *for* something, are implemented in the theoretical structure (Upmeier zu Belzen & Krüger, 2010).

In contrast to these scientific conceptions, most students refer to concrete objects such as replicas when thinking about the term 'model', or describe differences between model and original regarding the color, shape, dimension, or material (e.g. Grosslight et al., 1991; Trier & Upmeier zu Belzen, 2009). Another student conception is that there are multiple models to one and the same original because of different views of the same entity, such as different angles or inside/outside. It is rather rare that students justify different models by noticing different hypotheses (Grosslight et al., 1991). Teachers show similar conceptions of models (e.g. Crawford & Cullin, 2004, 2005; Van Driel & Verloop, 2002). A common conception is that '*only one, "the correct model", is possible for a particular phenomenon*' (Justi & Gilbert, 2003, p. 1375).

2.2 Concept of competence

The concept of competence is based on the definition of Weinert (2001), who understands competence '*as referring to combinations of those cognitive, motivational, moral, and social skills available to (or potentially learnable by) a person...that underlie the successful mastery through appropriate understanding and actions of a range of demands, tasks, problems, and goals*' (p. 2433). This study focuses on the cognitive part of Weinert's definition.

2.3 Model competence

2.3.1 Theoretical structure of model competence

Model competence for the school subject biology is understood as the ability (Upmeier zu Belzen & Krüger, 2010):

- to gain purposeful new insights into biological topics using models,
- to judge models in relation to their purpose, and
- to reflect on the epistemological process using models.

Upmeier zu Belzen and Krüger (2010) developed a theoretical structure based on empirical studies (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2002, 2003). It entails two cognitive dimensions of model competence: 1. ‘knowledge about models’ (Table 1) with the aspects ‘nature of models’ and ‘multiple models’ and 2. ‘modeling’ with the aspects ‘purpose of models’, ‘testing models’, and ‘changing models’. Each aspect is further divided into three levels which are based on Mahr’s (2009) three perspectives: the perspective on the model object itself, the creation of models, and the application of models. These levels do not describe developmental levels, as the research on this aspect is not yet complete.

2.3.2 Dimension ‘knowledge about models’

This article focuses on the dimension ‘knowledge about models’ (Table 1). This dimension describes individual cognitive concepts of models and deals with aspects of the nature of science. For the aspect ‘nature of models’, students compare the model with the original and comment on the extent to which the model is comparable with the original. Three positions are differentiated: a model is understood as a replication, as an idealized representation, or as a theoretical reconstruction (Upmeier zu Belzen & Krüger, 2010). The aspect ‘nature of models’ focuses on the perspective concerning the creation of models (Mahr, 2009). The aspect ‘multiple models’ refers to one and the same original being represented by different models. Students give varying explanations for these different models: they justify the presence of several models for one original by describing differences between the shown model objects, by arguing that the original allows the building of different models, or by noticing different hypotheses (Upmeier zu Belzen & Krüger, 2010). For the aspect ‘multiple models’, the levels describe the three perspectives of Mahr (2009).

Table 1
Structure of the dimension ‘knowledge about models’
(Upmeier zu Belzen & Krüger, 2010)

Level	I	II	III
Aspect			
Nature of models	replication of the original	idealized representation of the original	theoretical reconstruction of the original
Multiple models	differences between different model objects	the original allows the building of different models	different hypotheses about the original

3. Aims of the research project

The aims of our research project are to operationalize and evaluate the theoretical structure using open-ended tasks. Students' conceptions of models and modeling, and their impact on model competence, were determined during interviews (Trier & Upmeier zu Belzen, 2009). These served as background information for designing tasks. Compared to interviews, open-ended tasks are suitable for evaluating the theoretical structure of a larger sample in both (qualitative and quantitative) ways. Terzer, Krüger, and Upmeier zu Belzen (2009) realized these aims using multiple-choice, Krell and Krüger (2010) using forced-choice items, and Hänsch and Upmeier zu Belzen (2010) using hands-on assessment. These different approaches make it possible to cover different facets of model competence concerning thinking in, and handling of models.

4. Research questions

The first investigation was conducted to evaluate fifteen designed tasks:

To what extent do the student answers and comments indicate difficulties in understanding the tasks?

To what extent do the developed tasks enable students to answer at all three levels of the theoretical structure?

The tested and selected tasks were used to evaluate the theoretical structure:

To what extent are the student answers consistent with the theoretical structure?

5. Research design and method

5.1 Design of the open-ended tasks

For the first study, five open-ended tasks were designed for each aspect based on the theoretical structure. Each task consists of a short task context, a visualized biological model, and a standardized instruction (Table 2). The visualized figures represent functional, structural, and/or abstract models. The tasks were designed such that it should be possible to give answers at all levels.

Table 2
Standardized instructions for each aspect used in the first study

Dimension	Aspect	Instruction
Knowledge about models	Nature of models	Explain to what extent the model is comparable to the original.
	Multiple models	Explain why there are different models.

For the statistical analyses, the task pool needed to consist of three tasks for each aspect. These were selected based on whether the students understood the task context and if it was possible to give answers at all three levels (cf. 5.4, 6.1, 6.2). Three selected tasks of the aspect 'nature of models' were optimized and tested in a second investigation.

For the aspect 'multiple models', we could not continue to use either of the developed tasks and five tasks had to be redesigned, tested in a second study, and reduced to three tasks based on the above-mentioned criteria. Figures 1 and 2 exemplify tasks for both aspects with the above-described task structure, but with optimized instructions.

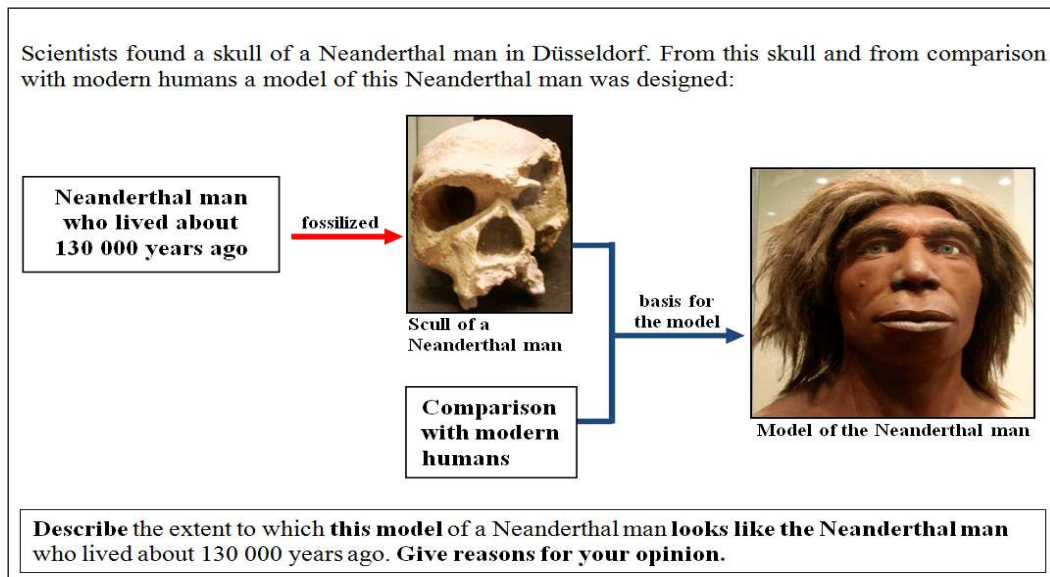


Figure 1

An optimized task for the aspect 'nature of models' used in the second study

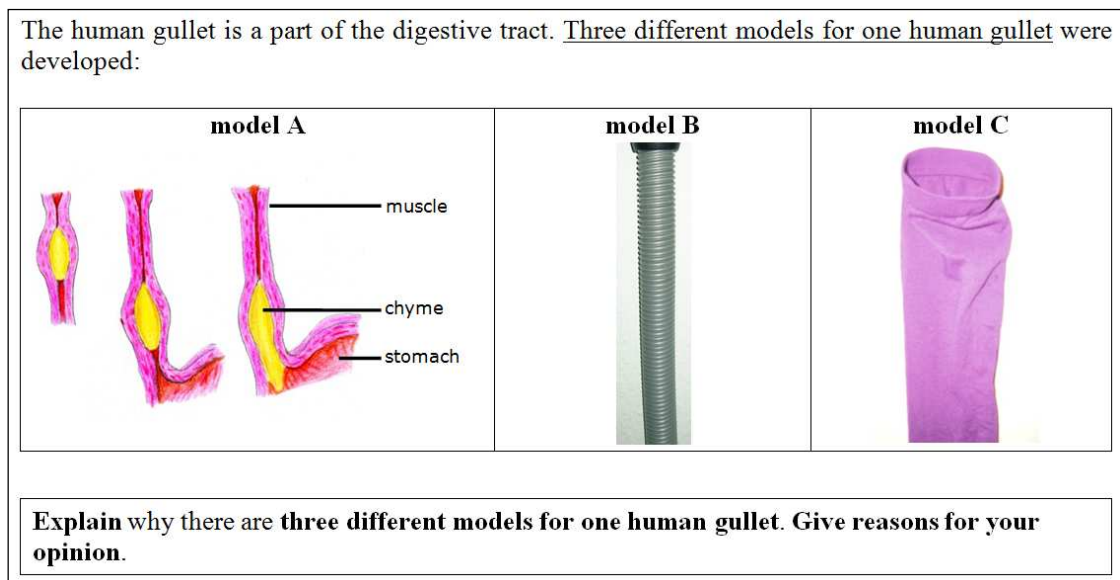


Figure 2

A redesigned task for the aspect 'multiple models' used in the second study

5.2 Sample

The first study with five tasks on each aspect was conducted with 406 7th and 10th-grade students (Realschule and Gymnasium, 12-17 years old). For the second study, three selected and optimized (nature of models) and five redesigned (multiple models) tasks were used and data were collected from a total of 506 8th, 9th and 10th-grade students (Realschule and Gymnasium, 13-18 years old). To obtain student answers at all three levels, different types of secondary schools and grades were chosen in both investigations. Since one student could not answer all tasks, the task pool was distributed in ten (first study) and five (second study) test booklets (multi-matrix design: Colbourn & Dinitz, 1996). Due to a test manual containing information about the research project, task format, and testing procedure, both investigations were standardized. In both studies, the students had 45 minutes to complete the test booklet. In the first study, each task was answered by about 100 students and in the second study, by about 130 (nature of models) or 70 (multiple models) students. Tasks relating to the same context were only compared with respect to students' responses if they were used in both studies.

5.3 Data processing

The data were analyzed by qualitative content analysis according to Mayring (2003), assisted by MAXQDA. Manual digitization and paraphrasing were used to obtain students' model competence. In an inductive approach, the student answers were summarized and categories were generated describing different students' conceptions of models. This categorization enabled a detailed description of each aspect. To evaluate the assigned student answers, additional coders were consulted. The assigned student answers from two coders were compared and differences were discussed until a consensus was reached (Gropengießer, 2001). Subsequently, the categories were assigned to the three levels of the theoretical structure (deductive approach).

5.4 Evaluation of the tasks

To determine whether students understand the tasks, a protocol was written during the investigation in which all upcoming questions and answers given by the test administrator were noted. In addition, the students were asked to make a note if they had problems understanding the text or figures. The level of correspondence between the instruction and the student answer is also a valuable indication of students' understanding of the tasks. In addition, the student answers were used to evaluate the extent to which it is possible to give answers at all three levels. If a student answered on different levels, his/her highest level was noted.

5.5 Consistency of the student answers with the theoretical structure

The student answers to the selected tasks were used to give some indication of the content evaluation of the theoretical structure. Therefore, the generated categories describing different students' conceptions were compared with the content of the theoretical structure.

6. Findings

6.1 Students' understanding of the designed tasks

6.1.1 Nature of models

The analysis of the student answers and comments showed a few difficulties in understanding certain parts of the task: a problem was caused by the term 'comparable' in the instruction: 'Explain to what extent the model is comparable to the original.' Some students used this term to answer the question, but they did not compare, or only insufficiently compared the model with the original: '*The mouth of the Neanderthal man is comparable with the mouth of the model*' (F1:82). The improved instruction (shown in Figure 1) led to more well-founded statements, for instance in the task 'Neanderthal man', there was an increase from 78% to 86% well-founded statements.

In addition, difficulties with certain contexts, such as DNA structure, were encountered. For this specific task, only 60% (n=102) of the student answers corresponded to the instruction and 17% of the students did not even answer this task. In contrast, only 8% (n=101) of the students did not answer the task 'Neanderthal man'. Therefore, those contexts were not used in subsequent studies.

6.1.2 Multiple models

Due to the imprecise and more open formulation of the instruction 'Explain why there are different models.' in the first study, 25% of the students (126 out of 510) justified the presence of different models with different originals. For this reason, 'one original' was referred to in the context and the instruction, as shown in Figure 2. The additional words were underlined in the context and highlighted in bold letters in the instruction. Despite these changes, 12% of the students (25 out of 204) in the second study still justified the presence of multiple models by different originals.

The data of the first study also revealed that students sometimes give short or no reasons for their opinion: '*There are two models in order to compare them with each other*' (F1:467). These reasons are important for a distinct assignment to the three levels of the theoretical structure. Therefore, the sentence 'Give reasons for your opinion.' was added to the instruction, to ensure well-founded statements. The optimized tasks (Figure 2) show pictures of models, making it easier to clearly identify more visual differences between the models such as different colors, different materials, or different perspectives.

As previously mentioned, students had difficulties understanding certain contexts, such as DNA replication. For this specific task, only 48% (n=102) of the student answers corresponded to the instruction; 16% of the students did not provide answers to this task. Contexts with such problems were not used further. In contrast, only 7% (5 out of 73) of the students did not give an answer to the redesigned task 'gullet'.

6.2 Student answers at all three levels

Most of the student answers to the aspects 'nature of models' (Table 3) and 'multiple models' (Table 4) could be assigned to the three levels of the theoretical structure. The developed tasks enabled students to give answers at their individual level.

6.2.1 Nature of models

Table 3 shows several prominent student answers for each level concerning the aspect ‘nature of models’. Students at level I are characterized by an understanding of models as an exact copy or an almost exact replication of the original: ‘*The model shows how a Neanderthal man looked like*’ (F1:476). In this context, some students show great confidence in scientific work and their technical tools: ‘*The figure of the Neanderthal man matches the truth because this picture was developed by experts*’ (F2:36); ‘*Due to recent visualizing techniques, a copy of an extinct species can be developed very easily*’ (F2:70). If students mentioned uncertainties concerning the appearance or noticed differences between the shown model and the original, they justified this either as a poor replication or by mentioning differences between their own ideas about the original and the model: ‘*The model does not look as primitive as the Neanderthal man should be*’ (F1:451).

Students at level II doubt that the model can be seen as an exact replication of the original. They realize that it is not necessary to duplicate everything of the original and that a model only shows extracts: ‘*The Neanderthal man looks more or less like the shown model. The height, length, and the size of the jawbones are right...However, what the hair looks like cannot be said*’ (F2:13).

At level III, students question the entire appearance of the model object and recognize the model as a theoretical reconstruction: ‘*The model is only a reconstruction or an idea which cannot match reality*’ (F2:116).

A comparison of the percentage distribution of student answers shows striking differences between the three levels. The majority of the students (50%/68% for first and second test, respectively) could be assigned to level I, followed by level II (16%/15%). Only a few students gave answers at level III (4%/3%).

Table 3
Prominent student answers for the aspect ‘nature of models’ for each level

Level	Frequency of student answers		Prominent student answers (student number)
	in absolute values and %, first test: (n=509)	in absolute values and %, second test: (n=392)	
I	253 (50%)	266 (68%)	<i>I think that the Neanderthal man does not look like the shown model because the model looks too fancy and the Neanderthal man should have a different look than that (F1:505)</i>
II	80 (16%)	59 (15%)	<i>I think that the model does not present the exact appearance of the Neanderthal man. However, it is possible to present an approximate appearance (F1:408)</i>
III	21 (4%)	12 (3%)	<i>The shown model is just a possibility of how a Neanderthal man could have looked. The bone structure has been recreated. ... However, they cannot be a 100 percent sure about that (F1:156)</i>

6.2.2 Multiple models

Table 4 exemplifies prominent student answers for each level of the aspect ‘multiple models’. Students at level I express their justification for the presence of multiple models to one and the same original by describing the model objects. They compare the model objects with each other and argue with different visualizations, materials, or material properties: *‘Model B and C are different by their appearance. Model B is rather inflexible, whereas model C is really flexible. Model A is a drawing’* (F2:64). An explicit reference to the original is not made. In fact, student F2:117 mentions the original *‘I think only model A looks like a human gullet,’* however, does not see the model as the result of a determined observation of the original (model of something).

Students that achieved level II justify the presence of multiple models to one and the same original by different key thematic aspects focused on by the modelers during the creation of the models: *‘There are different models because you can turn your attention to different characteristics of the gullet. Model A shows muscle movements, model B the structure, and model C shows that the gullet is not completely inflexible. To show all these characteristics in one model is almost impossible’* (F2:61). Here, an explicit reference to the original has been made.

At level III, students justify different models to one and the same original with different opinions of, or different hypotheses about the original, as exemplified in the following: *‘Because there are different theories and opinions of the human gullet, there are different models. The scientists could also have different opinions’* (F2:5).

Table 4 also shows the percentage distributions of the student answers. Most student answers in the first study (28%) could be assigned to level I, followed by level II (19%), whereas in the second study most student answers were given at level II (34%), followed by level I (27%). Level III (5%/15%) represents the lowest share in both studies.

6.3 Consistency of the student answers with the theoretical structure

For the aspect ‘nature of models’, the categories generated from student answers are consistent with the theoretical structure. For the aspect ‘multiple models’, the student answers show an additional perspective. As previously mentioned (cf. 6.1.2), 12% (n=204) of the students still justify the presence of multiple models with different originals, for instance: *‘There is only ONE heart. Why do we need three models?’* (F2:132); *‘There are different models because there are different types of gullets. Additionally, the shape of the gullet can change after some years’* (F2:352).

A further finding is that some students have alternative understandings of the term ‘model’. Figures that are clearly marked as models in the tasks are not understood as such. The alternative understandings of models are exemplified by the following statements: *‘...because A is a real model, however, B is only a drawing’* (F1:472) or *‘Model B is a hose or a tube and model C looks like a sock’* (F2:117).

Table 4

Prominent student answers for the aspect ‘multiple models’ for each level

Level	Frequency of student answers		Prominent student answers (student number)
	in absolute values and %, first test: (n=510)	in absolute values and %, second test: (n=204)	
I	145 (28%)	56 (27%)	<i>Model B is firm and not so flexible, whereas model C is very flexible (F2:118)</i>
II	98 (19%)	70 (34%)	<i>Model A represents the gullet with its muscles and stomach. Model B shall show how the tube with its muscles (really) looks from the outside. With model C you can see how the gullet adjusts to its content (F2:3)</i>
III	28 (5%)	30 (15%)	<i>There are always different opinions of biological processes or structures. And one person could...probably imagine the gullet as something inflexible and not flexible. And some other person imagines the gullet as something really flexible... (F2:49)</i>

7. Summary and discussion

The results of the second investigation show that students understand the selected tasks. Initial problems concerning the instruction and the associated insufficient explanations could be solved.

The majority of the student answers for the aspects ‘nature of models’ and ‘multiple models’ could be assigned to the theoretical structure (Upmeier zu Belzen & Krüger, 2010). The percentage distribution of the student answers for the aspect ‘nature of models’ (Table 3) revealed that most students perceive models as a replication (level I) of the original. Although the chosen models, such as the model of the Neanderthal man, do not allow an obvious reference to the original, only a few students think of a theoretical reconstruction (level III). These results correspond with the study by Grosslight et al. (1991), who determined that students refer to concrete objects rather than to representations of ideas or abstract entities when thinking about the term ‘model’.

The percentage distribution of the student answers for the aspect ‘multiple models’ (Table 4) in the first investigation demonstrates that most students explain the presence of multiple models by differences between the model objects in terms of color, shape, dimension, and material (level I). In the second study, most students justify the presence of multiple models by different thematic key aspects that modelers focus on during the creation of the models (level II). The conception of having different models because of different hypotheses about the original is rare in both studies (level III). Grosslight et al. (1991) note similar findings concerning levels I and II, showing that a prominent conception of students is that scientists could have different views of the same entities, such as different angles or inside/outside.

One possible reason for the prominent conceptions in both aspects of the theoretical structure might be a more frequent use of models as a substitute for the original or as a medium for transmitting information in biology lessons (e.g. Crawford & Cullin, 2004; Van Driel & Verloop, 2002).

For the aspect 'nature of models', the student answers are consistent with the theoretical structure. Referring to the aspect 'multiple models', the student answers show an additional facet: students justify the presence of multiple models with different originals even though there is a distinct reference to one original in the context and instruction of the redesigned tasks. Similar findings on the uniqueness of models are given by Justi and Gilbert (2003). Since the aspect 'multiple models' demands the conception of multiple models referring to one and the same original (Upmeyer zu Belzen & Krüger, 2010), the facet of uniqueness is missing in the theoretical structure of model competence. Therefore, this perspective needs to be added to the aspect 'multiple models' as an additional perspective. However, to ensure that students have the conception of uniqueness of models, the following task has to precede the open-ended tasks: the students have to decide between the choices of 'one model to one original' or 'different models to one original'.

A further finding is that some students have alternative conceptions of the term 'model', for instance, they do not perceive drawings or analogies as models. According to Mahr (2009), everything can be perceived and function as a model. In that context, the theoretical structure does not need to be changed. However, there is a need to gain access to this understanding by asking students in a separate questionnaire.

8. Prospects

The findings of the second study concerning the uniqueness of models will be integrated and tested in further investigations. In spring 2011, the subsequent study will be distributed among a larger sample of students ($N \approx 1000$, grades 7-10) in order to evaluate the indications gained in the described studies and the theoretical structure of model competence. In addition to open-ended items, multiple-choice items (Terzer et al., 2009; funded by BMBF) and forced-choice items (Krell & Krüger, 2010) will be used. This study will allow a comparison of the findings of all task formats.

REFERENCES

- Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Schümer, G., Stanat, P., Tillmann, K.-J., & Weiß, M. (2001). PISA 2000. Zusammenfassung zentraler Befunde. Berlin: Max-Planck-Institut für Bildungsforschung. Retrieved from <http://www.mpibberlin.mpg.de/pisa/ergebnisse.pdf>.
- Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22(9), 1041-1053.
- Colbourn, J. C., & Dinitz, H. D. (1996). *The CRC handbook of combinatorial designs*. New York: CRC Press.
- Crawford, B. A., & Cullin, M. J. (2004). Supporting prospective teachers' conceptions of modelling in science. *International Journal of Science Education*, 26(11), 1379-1401.

- Crawford, B. A., & Cullin, M. J. (2005). Dynamic assessments of preservice teachers' knowledge of models and modelling. In K. Boersma, M. Goedhart, O. de Jong & H. Eijkelhof (Eds.), *Research and the quality of education* (pp. 309-323). Dordrecht: Springer.
- Gilbert, S. W. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, 28(1), 73-78.
- Gropengießer, H. (2001). *Didaktische Rekonstruktion des »Sehens«*. Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung. Oldenburg: ZpB Zentrum für pädagogische Berufspraxis.
- Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28(9), 799-822.
- Hänsch, J., & Upmeier zu Belzen, A. (2010). Hands-on assessment zur Diagnose von Modellkompetenz im Biologieunterricht. In S. Nitz, M. Schreiber & K. Münchhoff (Eds.), *Tagungsband der 12. Frühjahrsschule der Fachsektion Biologiedidaktik im VBIO 2010* (pp. 102-103). Neumünster: Universitätsdruckerei Christian-Albrechts-Universität zu Kiel.
- Justi, R. S., & Gilbert, J. K. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. *International Journal of Science Education*, 24(4), 369-387.
- Justi, R. S., & Gilbert, J. K. (2003). Teachers' view on the nature of models. *International Journal of Science Education*, 25(11), 1369-1386.
- Kattmann, U. (2006). Modelle. In H. Gropengießer & U. Kattmann (Eds.), *Fachdidaktik Biologie*. Köln: Aulis Deubner.
- Krell, M., & Krüger, D. (2010). Deduktive Itemkonstruktion zur Diagnose von Modellkompetenz – Ergebnisse des Pretests der Items. In S. Nitz, M. Schreiber & K. Münchhoff (Eds.), *Tagungsband der 12. Frühjahrsschule der Fachsektion Biologiedidaktik im VBIO 2010* (pp. 96-97). Neumünster: Universitätsdruckerei Christian-Albrechts-Universität zu Kiel.
- Mahr, B. (2009). Information science and the logic of models. *Software and System Modeling*, 8(3), 365-383.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse*. Weinheim: Beltz UTB.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J., & Schiefele, U. (2004). PISA 2003. Ergebnisse des internationalen Vergleichs. Zusammenfassung. Retrieved from <http://www.pisa.ipn.uni-kiel.de/Ergebnisse_PISA_2003.pdf>.
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Wien und New York: Springer.
- Terzer, E., Krüger, D., & Upmeier zu Belzen, A. (2009). Diagnose von Modellkompetenz im Biologieunterricht – Validierung eines Kompetenzmodells mit Multiple Choice-Aufgaben. In U. Harms, F. X. Bogner, D. Graf, H. Gropengießer, D. Krüger, J. Mayer, B. Neuhaus, H. Prechtel, A. Sandmann & A. Upmeier zu Belzen (Eds.), *Referierter Tagungsband der Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO* (pp. 140-141). Kiel: Breitschuh & Kock GmbH.

- Trier, U., & Upmeyer zu Belzen, A. (2009). Die Wissenschaftler nutzen Modelle, um etwas Neues zu entdecken, und in der Schule lernt man einfach nur, dass es so ist. – Schülervorstellungen zu Modellen. In D. Krüger, A. Upmeyer zu Belzen, S. Hof, K. Kremer & J. Mayer (Eds.), *Erkenntnisweg Biologiedidaktik* 8 (pp. 23-38). Gießen/Marburg: Universitätsdruckerei Kassel.
- Upmeyer zu Belzen, A., & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *ZfDN*, 16, 41-57.
- Van Driel, J. H., & Verloop, N. (2002). Experienced teachers' knowledge of teaching and learning of models and modelling in science education. *International Journal of Science Education*, 24 (12), 1255-1272.
- Weinert, F. E. (2001). Competencies and key competencies: Educational perspective. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences*, vol. 4 (pp. 2433-2436). Amsterdam u. a.: Elsevier.

Beitrag 2

Grünkorn, J. & Krüger, D. (2012). Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells der Modellkompetenz. In U. Harms & F.X. Bogner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (Band 5, S. 9-27). Innsbruck: Studienverlag.

Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells zur Modellkompetenz

Zusammenfassung

Ziel dieses Forschungsprojekts ist die empirische Überprüfung eines Kompetenzmodells zur Modellkompetenz mit Aufgaben im offenen Antwortformat. Dieses Ziel kann mit Aufgaben erreicht werden, deren Antworten als Indikatoren für Modellkompetenz interpretiert werden können. Deshalb wurde nach der Entwicklung von 25 Aufgaben – für die Dimension *Modellbildung* – geprüft, inwiefern die gewählten Itemstämme für Schüler¹ plausibel sowie verständlich sind und inwiefern die Schülerantworten der jeweiligen Teilkompetenz und den drei Niveaus des Kompetenzmodells zugeordnet werden können. Dafür wurden die Aufgaben in zwei Teilstudien mit insgesamt $N=1011$ Schülern überprüft, ggf. optimiert und erneut getestet. Hinweise über die Verständlichkeit der Itemstämme bieten Schülerkommentare und -aussagen sowie Protokollnotizen aus den Testungen. Die Schüleraussagen wurden mit der qualitativen Inhaltsanalyse ausgewertet und ähnliche Schülerperspektiven über Modelle und Modellbildung zu Kategorien zusammengefasst. Die Ergebnisse der Untersuchungen zeigen, dass neun der entwickelten Aufgaben die oben genannten Kriterien erfüllen. Diese Aufgaben konnten somit Hinweise auf die inhaltliche Struktur des Kompetenzmodells geben: Bis auf zwei Schülerperspektiven zur Modellbildung (*Kein Testen von Modellen* und *Keine Änderung von Modellen*) werden mit dem Kompetenzmodell alle Perspektiven hinreichend beschrieben.

¹ Aus Gründen der besseren Lesbarkeit wird das maskuline Genus für beide Geschlechter verwendet, meint aber die weibliche und die männliche Form gleichermaßen.

Abstract

The objective of this research project is to evaluate a model of the theoretical structure of model competence using open-ended test items. To achieve this aim, test items are required that can be interpreted as indicators for model competence. For this purpose, 25 test items for the dimension modeling were developed. In two studies (N=1011), they were tested for understandability of the item stems and whether the student answers can be assigned to a certain aspect and the three levels of the theoretical structure. Student answers and comments as well as protocols written during the testing procedure give valuable information on the understandability of the item stems. The student answers were analyzed using qualitative content analysis and similar students' conceptions of models and modeling were summarized in categories. The analysis of the data showed that nine of the developed test items meet the above mentioned requirements. These items were used to give initial indications of the content evaluation of the theoretical structure: The majority of student answers are consistent with the theoretical structure. However, additional facets of the theoretical structure (no testing of models and no changing of models) could be described and need to be considered.

Einleitung

In den Naturwissenschaften spielt die Kompetenz, in Modellen denken und mit Modellen reflektiert umgehen zu können, eine wichtige Rolle (Harrison & Treagust, 2000). Eine Vielzahl von Studien (u. a. Grosslight, Unger, Jay & Smith, 1991; Trier & Upmeier zu Belzen, 2009) zeigt jedoch, dass Schüler die Rolle von Modellen für den wissenschaftlichen Erkenntnisprozess kaum wahrnehmen. Zu ähnlichen Ergebnissen kommen die PISA-Erhebungen von 2000 und 2003 (u. a. Prenzel et al., 2004), die zeigen, dass Schüler Schwierigkeiten haben, mentale Modelle in Problemlösesituationen heranzuziehen. Anhand eines Kompetenzmodells zur Modellkompetenz (Upmeier zu Belzen & Krüger, 2010) können Lehrkräfte die Förderung von Modellkompetenz strukturieren. Dieses Kompetenzmodell ist in zwei Dimensionen – den *Kenntnissen über Modelle* und der *Modellbildung* – und fünf Teilkompetenzen unterteilt und beschreibt den reflektierten Umgang mit Modellen. Um Lehrkräften gesicherte Hinweise zur Förderung geben zu können, ist es notwendig, das Kompetenzmodell zur Modellkompetenz empirisch zu überprüfen (Klieme et al., 2003). In diesem Forschungsprojekt wird dies mit Aufgaben im offenen Antwortformat realisiert. Damit Schüleraussagen, die zu Aufgaben im offenen Antwortformat gegeben werden, als Indikatoren der jeweiligen Teilkom-

petenz interpretiert werden können (Hartig, Frey & Jude, 2007), muss zunächst überprüft werden, inwiefern die gewählten Itemstämme für Schüler plausibel sowie verständlich sind und inwiefern die Schülerantworten der jeweiligen Teilkompetenz und den drei Niveaus des Kompetenzmodells zugeordnet werden können. Dieser Artikel fokussiert auf Ergebnisse zur Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat für die Dimension *Modellbildung*². Infolgedessen wird eine inhaltliche Überprüfung dieser Dimension mit Blick auf die empirische Gültigkeit des Kompetenzmodells vorgenommen.

Theoretischer Hintergrund

Modellbegriff

Mit dem Begriff *Modell* werden verschiedene Objekte wie beispielsweise Diagramme, Spielzeugautos, Realobjekte (z. B. *Drosophila melanogaster*) sowie auch Formeln verbunden. Trotz dieses breiten Begriffsverständnisses kann eine Gemeinsamkeit formuliert werden: Sie sind Träger von Wissen und Ideen (Mahr, 2009; Stachowiak, 1973).

Nach Mahr (2009) wird ein Modell durch zwei Betrachtungsweisen charakterisiert: die Herstellungsperspektive (*Modell von etwas*) und die Anwendungsperspektive (*Modell für etwas*). Mahr (2009) beschreibt, dass eine bestimmte Idee bzw. Vorstellung über etwas in einem Modellobjekt vergegenständlicht werden kann. Somit ist das Modellobjekt das Resultat eines Herstellungsprozesses (*Modell von etwas*). Diese Herstellung ist dabei an einen konkreten Zweck gebunden. So haben Watson und Crick (1953) ihre Ideen zur Doppelhelix-Struktur der DNA in einem 3D-Modellobjekt vergegenständlicht und konnten damit u. a. Vorhersagen über die Chargaff-Verhältnisse der Basenpaare treffen (Giere, Bickl & Mauldin, 2006). Ausgehend von dem im Modellobjekt repräsentierten Inhalt kann dieses Modellobjekt demnach auch eingesetzt und angewandt werden, um etwas über das Original zu erfahren bzw. seine Vorstellung über das Original zu verändern (*Modell für etwas*; Mahr, 2009).

Schülervorstellungen zur Modellbildung

Van Driel und Verloop (2002, S. 1147) konstatieren in ihrer Studie, dass angehende Lehrkräfte Modelle in der Schule eher als Mittel zur Veranschaulichung

2 Ergebnisse zur Dimension *Kenntnisse über Modelle* (Grünkorn, Upmeier zu Belzen & Krüger, 2011).

und weniger als Möglichkeit zur Vorhersage verstehen. Dieses Modellverständnis der Lehrkräfte beeinflusst nach Crawford und Cullin (2005, S. 310) dasjenige der Schüler. Vor diesem Hintergrund zeigen Interviewstudien von Grosslight et al. (1991) sowie Trier und Upmeier zu Belzen (2009), die Schüler offen zu Modellen und zur Modellbildung befragten, dass auch Schüler im Gegensatz zu Wissenschaftlern Modelle vorwiegend als Medien und weniger als Mittel zur Erkenntnisgewinnung verstehen und nutzen: Verbreitete Schülervorstellungen in diesen Interviewstudien zum *Zweck von Modellen* sind, Dinge oder Prozesse zu veranschaulichen bzw. zu erklären. Dabei betonen Schüler, dass Modelle vor allem als Hilfsmittel zum besseren Verständnis genutzt werden. Selten wird das Modell als Mittel zur Erkenntnisgewinnung herangezogen, mit dem Ideen bzw. Hypothesen untersucht werden können, um neue Erkenntnisse über das Original zu gewinnen (*purpose of models*; Grosslight et al., 1991, S. 806; *Zweck von Modellen*; Trier & Upmeier zu Belzen, 2009, S. 33). Auch bei der Frage, wie Modelle getestet werden können, überprüfen Schüler Modelle häufig aus der medialen Perspektive heraus, indem sie die Ähnlichkeit zwischen Modell und Original vergleichen (*purpose of models*; Grosslight et al., 1991, S. 807; *Testen von Modellen*; Trier & Upmeier zu Belzen, 2009, S. 33). Selten werden Modelle genutzt, um Ideen oder Hypothesen zu überprüfen (*purpose of models*; Grosslight et al., 1991, S. 807). Als Folge einer Testung von Modellen werden Änderungen am Modell vorgenommen. Ein Großteil der Schüler ist sich darin einig, dass Wissenschaftler Modelle ändern können (*changing a model*; Grosslight et al. 1991, S. 810). Vielfach wird das *Ändern von Modellen* mit neuen Erkenntnissen über das Original oder mit einer fehlerhaften Bauweise des Modells begründet. Nur wenige Schüler ziehen Modelle als Mittel zur Erkenntnisgewinnung in Betracht, so dass diese geändert werden, weil sich Hypothesen, die mit ihrer Hilfe aufgestellt und untersucht werden konnten, als nicht haltbar erwiesen haben (*changing a model*; Grosslight et al., 1991, S. 812f.; *Ändern von Modellen*; Trier & Upmeier zu Belzen, 2009, S. 33).

Theoretische Strukturierung von Modellkompetenz

Die theoretische Grundlage des Forschungsprojekts bildet ein Kompetenzmodell von Upmeier zu Belzen und Krüger (2010), das auf der Basis von empirischen Studien (z. B. Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003) sowie wissenschaftstheoretischer Literatur zu Modellen (u. a. Mahr, 2009) entwickelt wurde. Es differenziert zwei kognitive Dimensionen, die domänenübergreifendes und -spezifisches Wissen über Modelle enthalten (Leisner & Mikelskis, 2004): die Kenntnisse über Modelle mit den Teilkompetenzen *Eigenschaften von Modellen* und *Alternative Modelle* sowie die Modellbildung mit den Teilkompetenzen *Zweck*

von Modellen, Testen von Modellen und Ändern von Modellen (Tabelle 1). Jede Teilkompetenz wird wiederum in drei Reflexionsniveaus beschrieben. Dieser Artikel fasst Ergebnisse zur Dimension *Modellbildung* zusammen, weshalb diese im Folgenden näher beschrieben wird: Beim *Zweck von Modellen* gehen Schüler der Frage nach, welchen Nutzen Modelle haben können. Dabei nehmen sie das Modell zur Beschreibung von etwas (Niveau I) bzw. zur Erklärung von Zusammenhängen im Original (Niveau II) wahr oder nutzen es, um Vorhersagen über das Original zu treffen (Niveau III). Beim *Testen von Modellen* wird beschrieben, wie Modelle überprüft werden können. Entweder wird nur das Modellobjekt getestet, indem beispielsweise die Widerstandsfähigkeit des verwendeten Materials geprüft wird (Niveau I) bzw. das Modell wird mit dem Original verglichen (Niveau II) oder es werden Hypothesen über das Original mit dem Modell untersucht (Niveau III). Das Ergebnis einer Testung kann dazu führen, dass Modelle verändert werden müssen. Gründe für das *Ändern von Modellen* sind Mängel direkt am Modellobjekt z. B. Fehler am verwendeten Material (Niveau I), neue Erkenntnisse über das Original (Niveau II) oder eine Hypothese, die dem Modell zugrunde liegt, und die bei der Anwendung des Modells falsifiziert wurde (Niveau III).

Tabelle 1: Struktur und Niveaus der Dimension Modellbildung (Upmeyer zu Belzen & Krüger, 2010).

Teilkompetenz	Komplexität		
	Niveau I	Niveau II	Niveau III
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	Modell als Modell von etwas durch neue Erkenntnisse oder zusätzliche Perspektiven revidieren	Modell für etwas aufgrund falsifizierter Hypothesen revidieren

Fragestellungen und Hypothesen

Um ein Kompetenzmodell empirisch überprüfen zu können, müssen die dafür verwendeten Aufgaben das zugrunde liegende Konstrukt adäquat operationalisieren und eine Interpretation der Schülerantworten im Hinblick auf die Ausprägung der relevanten Kompetenz zulassen (Hartig et al., 2007). Für das Kompetenzmodell zur Modellkompetenz (Upmeyer zu Belzen & Krüger, 2010) bedeutet dies, dass die folgende zentrale Fragestellung vor der empirischen Überprüfung des Kompetenzmodells untersucht werden muss:

- (1) Inwiefern können die Schüleraussagen zu den entwickelten Aufgaben als Indikatoren für die entsprechende Teilkompetenz des Kompetenzmodells interpretiert werden?

Diese Fragestellung wird in Detailfragen untergliedert und dadurch überprüfbar:

- (1.1) Inwiefern sind die gewählten Itemstämme für Schüler plausibel und verständlich formuliert?
- (1.2) Inwiefern lassen sich die Schülerantworten auf die jeweilige Teilkompetenz beziehen?
- (1.3) Inwiefern bilden sich die drei Niveaus der jeweiligen Teilkompetenz in den Schülerantworten ab?

Das Kompetenzmodell (Upmeyer zu Belzen & Krüger, 2010) bildet die Grundlage der für jede Teilkompetenz entwickelten Aufgaben. Daher ist anzunehmen, dass die Aufgaben Modellkompetenz abbilden und ihre Bearbeitung vor diesem Hintergrund interpretiert werden kann. Die Aussagen zu den Aufgaben können dann als Indikatoren von Modellkompetenz angesehen werden, wenn die Aufgaben den Schülern der Zielgruppe (7.–10. Jahrgangsstufe) ein plausibles Modell bieten, die dazu gebotenen Informationen verständlich sind und die Fragestellungen Schülerantworten auslösen, die sich auf die jeweilige Teilkompetenz und deren Spektrum beziehen.

Ausgewählte evaluierte Aufgaben, die die oben genannten Kriterien erfüllen, wurden eingesetzt, um Hinweise auf die inhaltliche Struktur des Kompetenzmodells zur Modellkompetenz zu erhalten:

- (2) Inwieweit werden mit der theoretischen Struktur des Kompetenzmodells die Schülerperspektiven zu Modellen und zur Modellbildung umfassend beschrieben?

Das Kompetenzmodell von Upmeyer zu Belzen und Krüger (2010) basiert u. a. auf empirischen Studien von Crawford und Cullin (2005), Grosslight et al. (1991)

sowie Justi und Gilbert (2003). Daher wird erwartet, dass die identifizierten Schülerperspektiven zu Modellen und zur Modellbildung mit den dort beschriebenen Perspektiven übereinstimmen und entsprechend dem Kompetenzmodell zugeordnet werden können.

Untersuchungsdesign und Methodik

Entwicklung des Aufgabenpools

Mit Aufgaben im offenen Antwortformat kann die kognitive Facette von Modellkompetenz erfasst werden. Da Schülern mit einer offen gewählten Fragestellung die Möglichkeit gegeben wird, sich speziell auf das gezeigte Modell aber auch grundsätzlich über Modelle zu äußern, werden sowohl das konkrete (domänenspezifisch) als auch das abstrakte Wissen (domänenübergreifend) über Modelle in den Blick genommen (Leisner & Mikelskis, 2004). Über eine qualitative Evaluierung des Kompetenzmodells hinaus, die Trier und Upmeier zu Belzen (2009) in einer Interviewstudie durchgeführt haben, eignen sich Aufgaben im offenen Antwortformat für eine qualitative wie quantitative Überprüfung in einer größeren Stichprobe.

Für die Dimension *Modellbildung* wurden 25 Aufgaben im offenen Antwortformat – jeweils acht für die Teilkompetenzen *Zweck von Modellen* und *Testen von Modellen* und neun für die Teilkompetenz *Ändern von Modellen* – zu verschiedenen biologischen Inhalten entwickelt. Diese Aufgaben besitzen eine gemeinsame Struktur, die aus Itemstamm und Antwortformat besteht (Rost, 2004): Der Itemstamm liefert wesentliche Informationen zum behandelten Thema, zeigt das biologische Modell und schließt mit einer standardisierten Fragestellung (Tabelle 2). Er wurde so gewählt, dass ein Schüler je nach seiner Kompetenzausprägung plausibel auf allen drei Niveaus des Kompetenzmodells (Upmeier zu Belzen & Krüger, 2010) antworten kann. Für die Operationalisierung der Teilkompetenz *Zweck von Modellen* wurden vor allem Realobjekte als Modelle (z. B. Pflanzen in einer Schale als Wald-Modell) oder wie für die Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* technische Modelle (z. B. flugfähiges Libellen-Modell, flugfähiges Zanonía-Modell) genutzt. Diese lassen Perspektiven zu, die sowohl den Veranschaulichungsaspekt als auch die konkrete Anwendung der Modelle für Experimente verdeutlichen. Um den Einfluss von Lesekompetenz zu minimieren, ist der Itemstamm kurz gehalten.

Inwiefern der gewählte biologische Inhalt und die Modelle es ermöglichen, dass Schüler je nach ihrer Kompetenzausprägung plausibel auf allen drei Niveaus des Kompetenzmodells zur Modellkompetenz antworten, wurde vor der empi-

rischen Erprobung der Aufgaben mit Experten der Didaktik der Biologie diskutiert. Gleichmaßen wurde überprüft, ob die Inhalte fachlich richtig und für die Zielgruppe (7.–10. Jahrgangsstufe) u. a. auch sprachlich geeignet sind (Neuhaus & Braun, 2007). Da diese Aufgaben zudem in Berliner Realschulen mit z. T. hohem Migrationsanteil eingesetzt werden, wurden sie von einer DaZ-Expertin auf ihre Verständlichkeit und potentielle sprachliche Schwierigkeiten geprüft.

Tabelle 2: Ursprüngliche standardisierte Fragestellungen für die Teilkompetenzen.

Teilkompetenz	Fragestellung
Zweck von Modellen	Beschreibe, welchen Zweck dieses Modell hat!
Testen von Modellen	Begründe, wie man überprüfen kann, ob das dargestellte Modell tauglich ist bzw. seinen Zweck erfüllt!
Ändern von Modellen	Begründe, was alles dazu führen könnte, dass das Modell verändert wird!

Stichprobe und Untersuchungsablauf

In zwei Teilstudien wurden die entwickelten Aufgaben überprüft, ggf. optimiert und erneut getestet. Die erste Teilstudie umfasste $N=510$ Schüler und die zweite Teilstudie $N=501$ Schüler (Tabelle 3).

Tabelle 3: Angaben zur Stichprobe der Teilstudien 1 und 2.

Schulart	Jgst.	n	Geschlechterverteilung			Alter				
			♀	♂	k. A.	min.	max.	M	SD	k. A.
1. Teilstudie										
R	7	246	117	127	2	12	16	12.94	.725	3
GY	10	264	137	126	1	14	17	15.82	.557	1
2. Teilstudie										
R	7	101	48	51	2	12	16	13.15	.870	5
R	9	159	80	79	0	14	18	15.37	.753	1
GY	8	75	41	33	1	13	16	14.11	.481	0
GY	9	166	89	77	0	12	17	14.46	.824	2

Anmerkungen: R=Realschule, GY=Gymnasium, k. A.=keine Angaben

In den Studien wurden Berliner Schüler unterschiedlicher Jahrgangsstufen und Schultypen befragt, um das Antwortspektrum in der Breite zu erfassen. Da ein Schüler nicht alle Aufgaben beantworten konnte, wurden die Aufgaben in beiden Studien auf jeweils zehn Testhefte verteilt, wodurch pro Aufgabe ca. 100 (1. Teilstudie) bzw. ca. 60 (2. Teilstudie) Bearbeitungen vorliegen. Die Bearbeitungszeit des Fragebogens betrug bei beiden Teilstudien eine Schulstunde (45 Minuten). Die Durchführung wurde mit Blick auf ihre Objektivität in einem Testmanual, das Informationen zum Forschungsprojekt, zu Aufgabenformaten und zur Testdurchführung beinhaltete, standardisiert (Neuhaus & Braun, 2007).

Evaluierung der entwickelten Aufgaben

Die Schüler wurden in beiden Teilstudien dazu aufgefordert, die im Itemstamm gezeigten Modelle, Begriffe oder Abbildungen, die sie nicht verstanden, zu kennzeichnen und zu kommentieren. Darüber hinaus wurden während der Testungen alle aufkommenden Fragen der Schüler sowie die Antworten der Testleiter protokolliert. Zusätzlich bietet die Passung zwischen Frageintention und Antwort über die Fragen und Kommentare der Schüler hinaus einen Hinweis darauf, ob die Aufgaben für die Schüler verständlich formuliert sind. So wurde bei der Auswertung der Schüleraussagen darauf geachtet, ob Schüler die Informationen aus dem Itemstamm nur wiederholen und ob Antworten in der jeweiligen intendierten Teilkompetenz ausgelöst wurden bzw. über alle Schüleraussagen hinweg das Spektrum dieser Teilkompetenz bedient wurde. Falls Verständnisprobleme auftraten, wurden problemerezeugende Begriffe oder Abbildungen ersetzt oder Zusammenhänge anders dargestellt und diese Aufgaben erneut getestet. Zur Verbesserung der Aufgaben wurden sprachliche Vorschläge von Schülern bzw. Testleitern genutzt, um eine bessere Anbindung an die Sprache des Schülers zu gewährleisten (Neuhaus & Braun, 2007). Aufgaben, die nur von wenigen Schülern beantwortet wurden, wurden nicht weiter verwendet.

Auswertung der Schüleraussagen

Die Antworten der Schüler wurden nach den Schritten der qualitativen Inhaltsanalyse (Mayring, 2010) computergestützt mit der Software MAXQDA ausgewertet. Die Schüleraussagen wurden transkribiert und sprachlich geglättet, ohne den Sinn zu verändern. Ähnliche Schüleraussagen einer Teilkompetenz wurden zunächst zu Kategorien zusammengefasst und diese mit dem Kompetenzmodell der Modell-

kompetenz verglichen, um zu überprüfen, inwiefern alle Schülerperspektiven hinreichend beschrieben werden. Da die Reliabilität und Objektivität der Auswertung bei einem offenen Antwortformat von besonderer Bedeutung sind (Hartig et al., 2007), wurden zusätzlich mindestens zwei unabhängige Kodierer – Experten aus der Didaktik der Biologie – eingesetzt, die mithilfe eines Kodierleitfadens ebenfalls Zuordnungen vornahmen. Da in der ersten Teilstudie vordergründig das Ziel verfolgt wurde, Kategorien zu identifizieren und zu beschreiben, wurde eine Cohens-Kappa Berechnung nur für die Zuordnungen der zweiten Teilstudie vorgenommen (Bortz & Döring, 2002). Bei unterschiedlichen Zuordnungen zwischen den Kodierern wurde diskutiert, bis ein Konsens erreicht wurde. Schüler konnten gleichzeitig in verschiedenen Kategorien bzw. Niveaus antworten. Die fachliche Richtigkeit der Schüleraussage bzw. die Kompetenz bezüglich der adäquaten Planung einer Untersuchung wurden bei der Auswertung nicht beachtet.

Ergebnisse

Verständlichkeit des Itemstamms

Die Analysen der Schülerkommentare und -aussagen sowie der Protokollnotizen zeigten bei den Aufgaben für die Teilkompetenz *Zweck von Modellen*, dass Schüler die Informationen aus dem Itemstamm vollständig wiederholten bzw. der Itemstamm zu viele und vor allem lenkende Informationen enthielt (Tabelle 4). So antworteten vor der Überarbeitung einer Aufgabe in der ersten Teilstudie nur 2 % der Schüler auf Niveau III. Dies ließ sich durch Informationen im Itemstamm erklären, die Niveau I und II in den Vordergrund stellten (Tabelle 4). Nachdem diese Informationen stark gekürzt wurden, antworteten in einer zweiten Teilstudie 17 % der Schüler auf Niveau III. Auch für alle weiteren Itemstämme dieser Teilkompetenz wurde so vorgegangen, weshalb einige biologischen Inhalte (z. B. Enzyme, Schwimmblase) aufgrund der nun fehlenden Informationen im Itemstamm zu schwierig wurden und daher nicht weiter genutzt werden konnten.

Tabelle 4: Ursprüngliche und optimierte Version der Informationen im Itemstamm für die Aufgabe Wald der Teilkompetenz Zweck von Modellen.

Ursprünglich	Optimiert
Anna muss in ihrer Biologie-Klasse einen Vortrag zum Thema Wald halten. Dazu hat sie ein Stück Wald als Modell des Waldes in eine Schale gesetzt und ihren Mitschülern mitgebracht.	Im Foto siehst du das Modell eines Waldes.

Bei Aufgaben für die Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* traten hauptsächlich Probleme mit den Fragestellungen auf. Beim *Testen von Modellen* antworteten Schüler teilweise wie folgt: „Man kann sehen, wie sich die Libelle in bestimmten Situationen verhält, z. B. bei starkem Wind“ (F1:482). Da allerdings für Antworten in dieser Teilkompetenz vor allem die Art der Überprüfung und nicht nur der Zweck des Modells relevant ist, wurden die Begriffe *wie* und *überprüfen* durch eine zusätzliche Fettschreibung hervorgehoben. Bei einer Aufgabe, die diese Änderung beinhaltete, konnten in der zweiten Teilstudie 89 % (1. Teilstudie: 75 %) der Schüleraussagen den Kategorien dieser Teilkompetenz zugeordnet werden.

Schüler kommentierten häufig bei der Fragestellung der Teilkompetenz *Ändern von Modellen*, dass der Arbeitsauftrag nicht klar formuliert ist und sie den Teil *was alles dazu führen könnte* nicht verstehen (Tabelle 2). Da die durch den Testleiter gegebene Umformulierung zu *Nenne Gründe, warum dieses Modell verändert werden könnte!* besser verstanden wurde, wurde diese Fragestellung für die 2. Teilstudie übernommen. In dieser Teilstudie wurden nur reduziert Nachfragen gestellt und bei einer Aufgabe dieser Teilkompetenz konnten 79 % der Aussagen (1. Teilstudie: 71 %) den Kategorien zugeordnet werden.

Weiterhin gab es in allen Teilkompetenzen Aufgaben, deren biologische Inhalte für die Schüler zu schwierig waren und somit wenige Antworten in der jeweiligen Teilkompetenz auslösten. So wurde z. B. eine Aufgabe der Teilkompetenz *Zweck von Modellen* verworfen, weil neben 10 % nicht beantworteter Bearbeitungen zusätzlich bei 31 % der Schüler die Antwort nicht mit der Frage korrespondierte. Mit Blick auf alle drei Teilkompetenzen wurden aus den genannten Gründen 16 Aufgaben für folgende Studien nicht weiter verwendet.

Repräsentation der Teilkompetenzen

Die meisten identifizierten Kategorien, die verschiedene Perspektiven zur Modellbildung beschreiben, konnten den drei Niveaus der Teilkompetenzen *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen* des Kompetenzmodells zugeordnet werden.

Tabelle 5 präsentiert für die Teilkompetenz *Zweck von Modellen* ausgewählte Schüleraussagen pro Niveau. Schüler auf Niveau I (F2:299, Tabelle 5) sehen den Zweck von Modellen in der Darstellung des Sachverhaltes und beschreiben, was sie erkennen. Am Modell werden dabei keine Zusammenhänge zwischen verschiedenen Aspekten des Originals erkannt und beschrieben. In diesem Kontext bewerten Schüler die Modelle meist aus didaktischer Sicht und nehmen sie konkret als Hilfsmittel zum besseren Verständnis bei der Vermittlung wahr:

„Es soll [...] ein Hilfsmittel zur Veranschaulichung sein, um es sich besser vorzustellen“ (F2:163). Auch auf Niveau II können Schüler auf didaktischer Ebene antworten, also den Lerneffekt des Modells herausstellen (F2:197, Tabelle 5). Auf diesem Niveau zeichnen sich Schülerantworten vor allem darin aus, dass sie den Zweck von Modellen im Wiedererkennen, Nennen bzw. Erklären von Zusammenhängen zwischen verschiedenen Aspekten im Original sehen. Die Modelle werden genutzt, um bekannte Tatsachen wieder zu entdecken, aber nicht, um neue Erkenntnisse über das Original zu gewinnen. Letzteres ist ein Merkmal von Niveau III, bei dem Schüler eher eine wissenschaftliche Perspektive, weniger aber eine didaktische Sichtweise einnehmen. Der Zweck von Modellen wird auf diesem Niveau darin gesehen, Hypothesen bzw. Vermutungen über das Original zu überprüfen. Dabei werden Modelle als Instrumente zur Erforschung von Aspekten des Originals genutzt. Bei beiden Teilstudien verteilen sich die meisten Schülerantworten auf Niveau II (56 %/43 %), gefolgt von Niveau I (12 %/25 %). Nur wenige Schüler erkennen den Zweck von Modellen in der Überprüfung von Hypothesen (2 %/23 %).

Tabelle 5: Prozentuale Verteilung auf die drei Niveaus des Kompetenzmodells für die 1. (5 Aufgaben) und 2. Teilstudie (4 Aufgaben, $\kappa=0,84$) und prominente Schüleraussagen in der Teilkompetenz Zweck von Modellen. Die Werte repräsentieren das höchste erreichte Niveau (Niv.) der Schüler.

Niv.	1. Studie (n=513)	2. Studie (n=215)	Ankerbeispiele (Schülernummer)
I	12 %	25 %	„Dieses Waldmodell zeigt den Aufbau des Waldes.“ (F2:299)
II	56 %	43 %	„Dieses Modell soll den Schülern anschaulich machen, welche Pflanzen sich stärker/schwächer ausbilden. Dabei spielt Licht eine große Rolle. Wenn Pflanzen wenig Licht bekommen, werden sie nicht so gut wachsen.“ (F2:197)
III	2 %	23 %	„Dieses Waldmodell ist dazu da, um zu prüfen bzw. herauszufinden, welche Pflanzen zusammen wachsen können und ob die Erde dazu geeignet ist [...].“ (F2:187)

Tabelle 6 präsentiert die Ergebnisse für die Teilkompetenz *Testen von Modellen*. In dieser Teilkompetenz überprüfen Schüler auf Niveau I das Modell, indem sie bei-

spielsweise (F2:213, Tabelle 6) eine Materialprüfung bezüglich der Beweglichkeit, Stabilität und Elastizität durchführen. Sie prüfen, ob das Material des Modells der Prüfung standhält bzw. ob es dabei beschädigt werden kann. Weiterhin testen Schüler auf diesem Niveau, ob die Modelle die grundlegenden Voraussetzungen erfüllen, um z. B. weitere Testungen durchführen zu können. So formuliert Schüler F1:4: „Als erstes sollte man prüfen, ob das Modell auch fliegen kann. Sonst ist das ganze Modell meiner Meinung nach nicht so gut.“ Schülerantworten auf Niveau II (F2:206, Tabelle 6) nehmen Bezug zum Original und überprüfen das Modell, indem sie die Eigenschaften (Struktur und/oder Funktion) des Originals mit denen des Modells vergleichen. Auf Niveau III der Teilkompetenz *Testen von Modellen* prüfen Schüler mit dem Modell Hypothesen/Vermutungen über das Original, um neue Erkenntnisse über das Original zu gewinnen, und beschreiben, wie sie diese Vermutung mit dem Modell untersuchen können (F2:294, Tabelle 6). Nach den prozentualen Verteilungen der Schülerantworten antworten ein Großteil der Schüler auf Niveau II (34 %/53 %), gefolgt von Niveau I (10 %/20 %). Niveau III (5 %/9 %) ist am seltensten vertreten.

Tabelle 6: Prozentuale Verteilung auf die drei Niveaus des Kompetenzmodells für die 1. (5 Aufgaben) und 2. Teilstudie (4 Aufgaben, $\kappa=0,81$) und prominente Schüleraussagen in der Teilkompetenz Testen von Modellen. Die Werte repräsentieren das höchste erreichte Niveau (Niv.) der Schüler.

Niv.	1. Studie (n=509)	2. Studie (n=245)	Ankerbeispiele (Schülernummer)
I	10 %	20 %	„Man sollte das Modell von einem Baum herunterfallen lassen und sehen, ob es danach immer noch relativ unbeschadet und stabil ist.“ (F2:213)
II	34 %	53 %	„Man kann das Modell auf seine Maße [und] sein Gewicht prüfen. Die Struktur des Modells muss mit der des Originals übereinstimmen, sonst passt es nicht.“ (F2:206)
III	5 %	9 %	„Man kann das Modell mehrere Male aus einem Hochhaus fallen lassen und den Flug beobachten. Fliegt das Samenmodell mehrere Meter weiter weg, kann man daraus schlussfolgern, dass sich die Zanonía-Pflanze weit verbreiten kann.“ (F2:294)

Prominente Schüleraussagen der Teilkompetenz *Ändern von Modellen* werden in Tabelle 7 präsentiert. Schülerantworten auf Niveau I begründen (F1:553, Tabelle 7) ein Ändern des Modells mit Fehlern am Material. Dabei äußern Schüler u. a. Kritik an der Stabilität oder Elastizität des Materials. Ähnlich wie bei der Teilkompetenz *Testen von Modellen* begründen einige Schüler ein Ändern des Modells damit, dass das Modell nicht die grundlegenden Voraussetzungen erfüllt, die es braucht, um z. B. weitere Untersuchungen daran durchführen zu können: „Wenn das Modell zum Fliegen ist, und es das nicht macht, müssen die Wissenschaftler gründlich daran arbeiten“ (F2:496). Auf Niveau II beziehen sich Schüler in ihren Aussagen auf das Original und begründen ein Ändern des Modells entweder mit der mangelnden Passung des Modells mit dem Original oder mit neuen Erkenntnissen über das Original (F2:507, Tabelle 7). Wenden Schüler das Modell an, um ein Ändern des Modells zu begründen, antworten sie auf Niveau III. Dabei nutzen sie das Modell, um daran Daten zu gewinnen und um festzustellen, dass die Ergebnisse nicht mit den Ergebnissen am Original übereinstimmen. Ihre Hypothese über das Original, die implizit in den Antworten enthalten ist, wird verworfen und das Modell muss verändert werden (F1:148, Tabelle 7). In dieser Teilkompetenz antworten die meisten Schüler auf Niveau II (53 %/45 %). Deutlich weniger Schüler antworten auf Niveau I (10 %/13 %) und Niveau III (0,2 %/2 %) wird kaum erreicht.

Tabelle 7: Prozentuale Verteilung auf die drei Niveaus des Kompetenzmodells für die 1. (5 Aufgaben) und 2. Teilstudie (5 Aufgaben, $\kappa=0,81$) und prominente Schüleraussagen in der Teilkompetenz Ändern von Modellen. Die Werte repräsentieren das höchste erreichte Niveau (Niv.) der Schüler.

Niv.	1. Studie (n=510)	2. Studie (n=285)	Ankerbeispiele (Schülernummer)
I	10 %	13 %	„Schlechte Materialien könnten dazu führen, dass das Modell verändert wird.“ (F1:553)
II	53 %	45 %	„Wenn neue Informationen über die Libelle da sind, muss man sie einbauen.“ (F2:507)
III	0,2 %	2 %	„Da man mit einem nachgestellten Modell viel mehr Dinge überprüfen kann, ändern sich Modelle. Da bekommt man dann vielleicht bei einer Testung heraus, dass etwas nicht so wichtig war, wie man gedacht hat und dann muss man seine Idee und sein Modell ändern.“ (F1:148)

Überprüfung der inhaltlichen Struktur des Kompetenzmodells

Ausgewählte evaluierte Aufgaben konnten Hinweise auf die inhaltliche Struktur des Kompetenzmodells geben. Für die Teilkompetenz *Zweck von Modellen* werden im Kompetenzmodell alle Schülerperspektiven umfassend beschrieben. In den Schülerantworten der Teilkompetenz *Testen von Modellen* wurde eine Perspektive identifiziert, die den drei Niveaus des Kompetenzmodells nicht zugeordnet werden konnte. Obwohl den Schülern Modelle präsentiert wurden, mit denen eine Anwendung bzw. Testung möglich und plausibel ist, äußerten einige wenige Schüler (1 %) ihre Ratlosigkeit diesbezüglich, was an Modellen zu überprüfen ist. Beispielsweise antworten sie: „Warum soll man denn dieses Modell testen? Meiner Meinung nach ist dies unnötig.“ (F3:71); „Ich habe leider keine Ahnung, wie und warum man das testen soll“ (F2:185). Auch für die Teilkompetenz *Ändern von Modellen* wurde eine zusätzliche Kategorie identifiziert. Einige Schüler (4 %) beantworteten die Frage nach Gründen für ein Ändern folgendermaßen: „Keine Ahnung, warum soll sich das Modell verändern. Das macht doch keinen Sinn“ (F3:141). Solche Antworten zeigen, dass die Schüler keinen Grund darin sehen, Modelle zu ändern oder keine Ideen haben, warum Modelle verändert werden sollten.

Diskussion

Die Ergebnisse der Teilstudien zeigen, dass davon auszugehen ist, dass neun von 25 entwickelten Aufgaben Antworten zulassen, die als Indikatoren für Modellkompetenz interpretiert werden können (Hartig et al., 2007). Diese Aufgaben bieten ein plausibles Modell, verfügen über nachvollziehbare Informationen zum Thema sowie über verständliche Fragestellungen für Schüler der Zielgruppe (7.–10. Jahrgangsstufe, Realschule und Gymnasium) und lösen Schülerantworten aus, die sich auf die jeweilige Teilkompetenz und die drei Niveaus des Kompetenzmodells (Upmeier zu Belzen & Krüger, 2010) beziehen. Die vorgenommenen Änderungen an den Aufgaben könnten dazu beigetragen haben, dass die Aufgaben im offenen Antwortformat von den Schülern besser verstanden wurden, wodurch vermehrt Aussagen den Kategorien der jeweiligen Teilkompetenzen zugeordnet werden konnten. Die Werte von Cohens Kappa liegen bei den drei Teilkompetenzen der zweiten Teilstudie zwischen $\kappa=0,81$ und $\kappa=0,84$ und sind als gut einzuschätzen (Bortz & Döring, 2002, S. 277). Die Verteilungen der Schüleraussagen auf die drei Niveaus der Teilkompetenzen *Zweck von Modellen* (Tabelle 5), *Testen von Modellen* (Tabelle 6) und *Ändern von Modellen* (Tabelle 7)

sind mit den Befunden von Grosslight et al. (1991) sowie Trier und Upmeier zu Belzen (2009) vergleichbar: Ein überwiegender Teil der Schüler sieht den Zweck von Modellen darin, Objekte zu veranschaulichen und Zusammenhänge im Original zu erklären. Beim Testen von Modellen parallelisieren die Schüler vorrangig das Modell mit dem Original. Als Änderungsgründe werden die mangelnde Passung des Modells mit dem Original bzw. neue Erkenntnisse über das Original genannt. Schüler charakterisieren ein Modell somit mehrheitlich aus einer medialen Perspektive heraus. Die methodische Perspektive, bei der Modelle bei der Anwendung (Mahr, 2009) als Mittel zur Erkenntnisgewinnung angesehen werden, wird nur von wenigen Schülern erkannt (Grosslight et al., 1991; Trier & Upmeier zu Belzen, 2009). Geht man wie Crawford und Cullin (2005) davon aus, dass das Modellverständnis der Lehrkräfte das der Schüler beeinflusst, könnten diese prominenten Schülervorstellungen zur Modellbildung darauf hinweisen, dass Modelle als Mittel zur Voraussage und Generierung von Hypothesen selten im Unterricht eingesetzt werden und möglicherweise der Umgang mit und das Denken in Modellen kaum thematisiert werden (Van Driel & Verloop, 2002). Weiterhin sollte aber auch darüber reflektiert werden, inwieweit das Antwortverhalten der Schüler von dem gezeigten Modelltyp beeinflusst werden kann und dadurch entsprechende Ergebnisse erreicht werden. Diese Fragestellung wird in einer Studie von Krell, Upmeier zu Belzen und Krüger (2012, einger.) untersucht. Für die Teilkompetenz *Zweck von Modellen* beschreibt die Struktur des Kompetenzmodells (Upmeier zu Belzen & Krüger, 2010) hinreichend alle auftretenden Schülerperspektiven. Für die Teilkompetenzen *Testen von Modellen* und *Ändern von Modellen* konnten Schülerantworten in Kategorien beschrieben werden, die nicht in der Struktur des Kompetenzmodells (Upmeier zu Belzen & Krüger, 2010) abgebildet sind. Bereits die Studie von Grosslight et al. (1991) weist auf die Perspektive *Keine Änderung des Modells* hin. Bezüglich der Kategorie *Kein Testen von Modellen* werden in den Studien von Crawford und Cullin (2005), Grosslight et al. (1991) sowie Justi und Gilbert (2003) keine Hinweise gegeben. Allerdings kommt eine Studie von Hänsch und Upmeier zu Belzen (2012), die mit Hands-On Aufgaben Modellkompetenz erfasst, zu ähnlichen Ergebnissen. Zur Absicherung dieses Befundes wurden zusätzlich Aufgaben im dichotomen Antwortformat entwickelt, die den Aufgaben im offenen Antwortformat vorgeschaltet werden. Hier müssen Schüler grundsätzlich Stellung beziehen, ob Modelle getestet bzw. geändert werden können. Wenn sich diese weiteren Perspektiven als bedeutsam erweisen, muss dies bei einer Förderung von Modellkompetenz entsprechend berücksichtigt werden.

Ausblick

Um die Hinweise aus den vorgestellten Teilstudien zu überprüfen und das Kompetenzmodell zur Modellkompetenz (Upmeier zu Belzen & Krüger, 2010) inhaltlich zu konkretisieren und zu überprüfen, werden in einer weiteren Untersuchung je Teilkompetenz drei überprüfte Aufgaben im offenen Antwortformat eingesetzt. Ergänzend werden für die Teilkompetenzen *Alternative Modelle* (Grünkorn et al., 2011), *Testen von Modellen* und *Ändern von Modellen* je drei Aufgaben im dichotomen Antwortformat vorgeschaltet. Zudem ist geplant, auf der Basis des entwickelten und überprüften Kategoriensystems, geschlossene Aufgabenformate zu konstruieren, mit denen Modellkompetenz diagnostiziert werden kann.

Dank

Wir bedanken uns bei der Deutschen Forschungsgesellschaft für die Unterstützung des Projekts und bei den Gutachtern für ihre wertvollen Anregungen zu diesem Manuskript.

Literaturverzeichnis

- Bortz, J., & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Crawford, B., & Cullin, M. (2005). Dynamic assessments of preservice teachers' knowledge of models and modelling. In K. Boersma, H. Eijkelhof, M. Goedhart, & O. Jong (Eds.), *Research and the Quality of Science Education* (pp. 309–323). Dordrecht: Springer.
- Giere, R. N., Bickl, J., & Mauldin, R. F. (2006). *Understanding scientific reasoning* (5. Aufl.). Toronto: Thomson Wadsworth.
- Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28(9), 799–822.
- Grünkorn, J., Upmeier zu Belzen, A., & Krüger, D. (2011). Design and test of open-ended tasks to evaluate a theoretical structure of model competence. In A. Yarden & G. S. Carvalho (Eds.), *Authenticity in Biology Education: Benefits and Challenges. A selection of papers presented at the VIIIth Conference of European Researchers in Didactics of Biology (ERIDOB) Braga, Portugal* (pp. 53–65). Braga Portugal: CIEC Universidade do Minho.

- Hänsch, J., & Upmeyer zu Belzen, A. (2012, März). *Hands-On Aufgaben zur Erfassung und Förderung von Modellkompetenz*. Vortrag auf der 14. Frühjahrsschule der Fachsektion Didaktik der Biologie, Bremen.
- Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, 22(9), 1011–1026.
- Hartig, J., Frey, A., & Jude, N. (2007). Validität. In H. Moosbrugger (Ed.), *Testtheorie und Fragebogenkonstruktion* (pp. 136–163). Heidelberg: Springer Medizin.
- Justi, R. S., & Gilbert, J. K. (2003). Teachers' views on the nature of models. *International Journal of Science Education*, 25(11), 1369–1386.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards*. Bonn: BMBF.
- Krell, M., Upmeyer zu Belzen, A., & Krüger, D. (2012). Analysis of Students' Understanding of Models and Modelling in Biology Education: Empirical Evaluation of a Theoretical Structure. Zur Veröffentlichung eingereicht.
- Leisner, A., & Mikelskis, H. F. (2004). Erwerb metakonzeptueller Kompetenz durch ein systematisches Lernen über Modelle. In A. Pilon (Hrsg.), *Zur Didaktik der Physik und Chemie. Beitragsband zur Tagung in Berlin 2003* (S. 120–122). Münster: LIT Verlag.
- Mahr, B. (2009). Information science and the logic of models. *Software and Systems Modeling*, 8, 365–383.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (11., aktualisierte und überarb. Aufl.). Weinheim: Beltz.
- Neuhaus, B., & Braun, E. (2007). Testkonstruktion und Testanalyse – praktische Tipps für empirisch arbeitende Didaktiker und Schulpraktiker. In H. Bayrhuber, D. Elster, D. Krüger & H. J. Vollmer (Hrsg.), *Forschungen zur Fachdidaktik: Bd. 9. Kompetenzentwicklung und Assessment* (S. 135–164). Innsbruck: Studienverlag.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J., & Schiefele, U. (Eds.). (2004). *PISA 2003. Ergebnisse des zweiten internationalen Vergleichs. Zusammenfassung*. Zugriff am 30. Oktober. 2001, von http://www.ipn.uni-kiel.de/pisa/Zusammenfassung_2003.pdf
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Wien: Springer.
- Trier, U., & Upmeyer zu Belzen, A. (2009). Die Wissenschaftler nutzen Modelle, um etwas Neues zu entdecken, und in der Schule lernt man einfach nur, dass es so ist. – Schülervorstellungen zu Modellen. In D. Krüger, A. Upmeyer zu Belzen, S. Hof, K. Kremer, & J. Mayer (Eds.), *Erkenntnisweg Biologiedidaktik 8* (pp. 23–37). Kassel: Universitätsdruckerei.

- Upmeyer zu Belzen, A., & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57.
- Van Driel, J. H., & Verloop, N. (2002). Experienced teacher's knowledge of teaching and learning of models and modelling in science education. *International Journal of Science Education*, 24(12), 1255–1277.
- Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids. A structure for desoxyribose nucleic acid. *Nature*, 171(4356), 737–738.

Beitrag 3

Grünkorn, J., Upmeyer zu Belzen, A. & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education*, 1-34. <http://dx.doi.org/10.1080/09500693.2013.873155>

Assessing Students' Understandings of Biological Models and their Use in Science to Evaluate a Theoretical Framework

Juliane Grünkorn^{a*}, Annette Upmeier zu Belzen^b and Dirk Krüger^c

^aDepartment of Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Germany;

^bBiology Education, Humboldt-Universität zu Berlin, Berlin, Germany; ^cBiology Education, Freie Universität Berlin, Berlin, Germany

Research in the field of students' understandings of models and their use in science describes different frameworks concerning these understandings. Currently, there is no conjoint framework that combines these structures and so far, no investigation has focused on whether it reflects students' understandings sufficiently (empirical evaluation). Therefore, the purpose of this article is to present the results of an empirical evaluation of a conjoint theoretical framework. The theoretical framework integrates relevant research findings and comprises five aspects which are subdivided into three levels each: nature of models, multiple models, purpose of models, testing, and changing models. The study was conducted with a sample of 1,177 seventh to tenth graders (aged 11 – 19 years) using open-ended items. The data were analysed by identifying students' understandings of models (nature of models and multiple models) and their use in science (purpose of models, testing, and changing models), and comparing as well as assigning them to the content of the theoretical framework. A comprehensive category system of students' understandings was thus developed. Regarding the empirical evaluation, the students' understandings of the nature and the purpose of models were sufficiently described by the theoretical framework. Concerning the understandings of multiple, testing, and changing models, additional initial understandings (only one model possible, no testing of models, and no change of models) need to be considered. This conjoint and now empirically tested framework for students' understandings can provide a common basis for future science education research. Furthermore, evidence-based indications can be provided for teachers and their instructional practice.

*Corresponding author. Department of Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Germany. Email: gruenkorn@dipf.de

Keywords: Students' understandings of models and their use in science; Theoretical framework; Competence; Assessment; Empirical evaluation

Introduction

Models and the process of modelling are considered key elements for the work of scientists but also for citizens' participation in social discourses and in decision-making processes in their everyday life (Odenbaugh, 2005; Oh & Oh, 2011). To understand and evaluate the work of scientists as well as their way of conceptualising phenomena and to participate in scientific discourses, it is necessary to learn and know about models and their use in science (Grosslight, Unger, Jay, & Smith, 1991). Thus, science educational standards place a considerable emphasis on models and their use in science and require students to be knowledgeable in these aspects (American Association for the Advancement of Science, 1993; National Research Council, 2000; Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005).

Based on the relevance of models as well as of the process of modelling and its reflection, research has focused on investigating students' (Grosslight et al., 1991; Schwarz et al., 2009; Schwarz & White, 2005; Treagust, Chittleborough, & Mamiala, 2002; Trier & Upmeier zu Belzen, 2009), teachers' (Crawford & Cullin, 2004, 2005; Justi & Gilbert, 2002, 2003; Van Driel & Verloop, 1999), and scientists' (Grosslight et al., 1991; Van Der Valk, Van Driel, & De Vos, 2007) understandings of models and their use in science. In addition to investigating learners' understandings, other scholars have drawn attention to students' handling of models (Louca, Zacharia, Michael, & Constantinou, 2011; Schwarz et al., 2009). As a result, various frameworks were developed. Some of these frameworks are used for the analysis of learners' understandings of models, yet they describe and structure aspects of models differently (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Treagust et al., 2002). Others provide analytical frameworks that combine learners' understandings and their handling of models (Schwarz et al., 2009). The critical issue of all these different frameworks lies in the absence of an overall empirical evaluation, i.e. whether these frameworks describe learners' understandings sufficiently (cf. Schwarz et al., 2009, p. 637).

Consequently, two research gaps can be identified and need to be addressed. First, research needs to provide a framework for analysing learners' understandings of models and their use in science. This framework ought to focus solely on learners' understandings, and bring together as well as integrate the different approaches to learners' understandings. Second, this developed framework needs to be empirically tested with a large sample to evaluate if learners' understandings are sufficiently reflected by the framework (cf. Klieme et al., 2007). Such a conjoint and empirically tested framework can be profoundly beneficial for science education research as well as for science education. For science education research, it could be used as a basis for future research studies. Studies that refer to the same framework allow for valid

comparisons between their results (cf. Hartig, 2008). In addition, this framework creates a central precondition for addressing the demand raised by Louca and Zacharia (2012) to investigate the relationships between learners' understandings of models, practical skills, and other components influencing students' processes of modelling. This demand could be realised by using frameworks for each component and by assessing these components separately. Furthermore, teachers can gain information about evidence-based indications for their instructional practices.

Regarding the first research gap, Upmeier zu Belzen and Kruger (2010) contribute to this area of research by combining and integrating different empirical studies (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003) and by consulting science-theoretical approaches (Giere, 2004; Mahr, 2009). This theory-based combination of empirical studies results in a theoretical framework for students' understandings of models and their use in science (Upmeier zu Belzen & Kruger, 2010). The purpose in this article is to target the second research gap and, therefore, show the results of an overall empirical evaluation of the theoretical framework. This issue is addressed by using open-ended items that are situated in different biological model contexts and by assessing a large sample representing students of different ages. The learners' understandings are investigated in the domain of biology as the understandings of models and their use in science in general can be understood as cognitive dispositions which are acquired by learning in a specific domain (cf. Klieme, Hartig, & Rauch, 2008).

Theoretical Background

Different Frameworks for Learners' Understandings of Models and their Use in Science

So far, all efforts concerning learners' understandings of models and their use in science have been focused on generating and describing different frameworks and investigating learners' understandings (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Louca et al., 2011; Schwarz et al., 2009). Grosslight et al.'s (1991) study is considered as fundamental to research on students' understandings of models and their use in science. They elicited five aspects (kinds of models, multiple models for the same thing, purpose of models, designing and creating models, and changing a model) and described three general levels of models and their use in science from the analysis of their interview study with high school students. Students at level 1 understand models as simple copies of reality with the purpose of matching the real thing in colour, shape, dimension, or material. At level 2, they realise that the construction of a model is linked to a specific purpose. Therefore, models are not seen as exact duplicates of reality. Students at level 3 are aware that the original (target, phenomenon, or object) is explained through examination and/or manipulation of the model. Here, the model is used as a method to test ideas and to draw conclusions on the original. Based on this characterisation, Grosslight et al. (1991) stated that the majority (67 %) of students were at level 1, followed by level 2 (12%). None of the students achieved level 3. Similar results have been

obtained in other studies showing that students reflect little on their thinking in models and that they are not aware of the role models play in epistemological processes (Schwarz & White, 2005; Trier & Upmeier zu Belzen, 2009).

Drawing on Grosslight et al.'s (1991) three general levels, Justi and Gilbert (2003) investigated the response patterns regarding models of 39 science teachers. However, they could not provide any support for these general levels. Moreover, they generated seven aspects (nature, use, entities, uniqueness, time, predictions, and accreditation) from the analysis of interviews. Contrary to Grosslight et al.'s (1991) structure, they subdivided the aspect of kinds of models into nature and entities and the aspect of purpose of models into use, predictions, and accreditation. Furthermore, Justi and Gilbert (2003) explicitly described the teacher understanding of only one model being possible for a phenomenon within the aspect of uniqueness (p. 1375) in their study. With regard to students' understandings of models, empirical evidence is needed to find out whether students endorse this kind of understanding as well.

A different framework is provided by Crawford and Cullin (2005) which they deduced from a review of relevant literature. They distinguished between five aspects (multiple models for the same thing, purpose of models, designing and creating models, validating/testing models, and changing a model) and described three to four levels (limited, pre-scientific, emerging scientific, and scientific) for each aspect. In contrast to Grosslight et al.'s (1991) aspect of kinds of models and Justi and Gilbert's (2003) aspects of nature and entities, they did not specify an aspect whereby the relationship between the model and the original is characterised. Besides, without an overall empirical evaluation with a large sample regarding structure and content, Crawford and Cullin (2005) used their framework as an analytical basis for tracking the understandings of models and modelling of eight teachers.

Other than the frameworks described above which refer to learners' understandings of models and their use in science, Louca et al. (2011) as well as Schwarz et al. (2009) rather focus on the handling of models and students' practical skills. For instance, Schwarz et al. (2009) developed a learning progression that integrated practical skills (called elements of the practice), learners' understandings of models (called meta-modelling knowledge), and communication aspects. These are further divided into four levels of performance. Schwarz et al. (2009) stated that an 'overall validation of the progression' (p. 637) is still missing. However, in their recent publication (Schwarz, Reiser, Acher, Kenyon, & Fortus, 2012) they suggest refinements of their learning progression after applying it in classroom since it was not possible to clearly assign students to one of the levels. Findings of this study (Schwarz et al., 2012) support an approach to define and assess these different components (understanding, practical skills, and social components of models and modelling) separately. It is thus possible to address the issue claimed by Louca and Zacharia (2012) to investigate and analyse the relationships between the understanding, practical skills, and other components influencing the process of modelling.

In summary, the described approaches show different frameworks for learners' understandings of models and their use in science (for a detailed discussion see Krell, Upmeier zu Belzen, & Kruger, 2013; Upmeier zu Belzen & Krüger, 2010),

while no conjoint framework has been developed to integrate the relevant results of these approaches. Furthermore, empirical evidence is needed to evaluate whether the learners' understandings of models and their use in science are sufficiently described. This article meets this demand by portraying a theoretical framework for students' understandings of models and their use in science and by evaluating this framework with a large sample and students of different ages.

Theoretical Framework for Learners' Understandings of Models and their Use in Science

The described frameworks (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003) and science-theoretical approaches (Giere, 2004; Mahr, 2009) are integrated into a theoretical framework (Upmeier zu Belzen & Kruger, 2010). This theoretical framework can be used as an analytical framework for assessing and investigating learners' understandings of models and their use in science. Unlike the approach pursued by Schwarz et al. (2009), this framework solely emphasises the cognitive component of models (cf. Klieme et al., 2008; Weinert, 2001). Such a focused approach is strongly in line with a concept of competence underlying international large-scale assessments (e.g., Programme for International Student Assessment). The concept of competencies relates to Klieme and Leutner's (2006) definition of competencies 'as context-specific cognitive dispositions that are acquired by learning and needed to successfully cope with certain situations or tasks in specific domains' (Klieme et al., 2008, p. 9). In this regard, the cognitive component is a basis for successfully solving demanding problems in various situations (cf. Fleischer, Koeppen, Kenk, Klieme, & Leutner, 2013; Klieme et al., 2008; Weinert, 2001). Notably, the authors support the idea that the understanding of models, the handling of models (Louca et al., 2011; Schwarz et al., 2009), as well as social components (Oh & Oh, 2011; Schwarz et al., 2009) are intertwined and thus cannot exist in isolation. Nevertheless, for the purpose of investigating the relationships between these components, a separate definition, framework, and assessment is highly beneficial.

In general, the theoretical framework refers to all models collectively without focusing on specific types of models (Boulter & Buckley, 2000). An essential basis for the theoretical framework is an understanding of models from a medial perspective as an illustration of something, for instance, an idea of the original (target, phenomenon) and from a methodical perspective as an instrument for something, i.e. for testing ideas and drawing conclusions on the original (Gilbert, 1991; Mahr, 2009; Oh & Oh, 2011).

Deduction of the Levels of the Theoretical Framework

Grosslight et al. (1991) generated three general levels of models and their use that are closely tied to a person's epistemological views on science ranging from a naive-realistic view to constructivist understandings of the epistemology of science (Hofer & Pintrich, 1997). Justi and Gilbert's (2003) study, however, provided no support for Grosslight et al.'s (1991) levels but supported the idea of defining three levels

for each aspect. In addition, Crawford and Cullin (2005) described at least three levels for each aspect. As a result, Upmeier zu Belzen and Kruger (2010) defined three levels for each aspect (cf. Deduction of the aspects of the theoretical framework) of the theoretical framework. These three levels reflect the epistemological viewpoint of models as products as well as methods of science (Gilbert, 1991; Mahr, 2009): On a basic level I, only the representational model is being considered without focusing on the original (called model object, Mahr, 2009). Perceiving the model as a medium of something and focusing on the creation process describes level II (called model of something, Mahr, 2009). Both perspectives are complemented by a methodological view of models (level III) in their application comprehending their use in science to test and draw conclusions on the original (called model for something, Mahr, 2009). Recently, studies investigating the three levels of this theoretical framework using multiple choice and forced-choice items provide empirical evidence that these levels reflect an increasing degree of difficulty (Krell, 2012; Terzer, 2013).

Deduction of the Aspects of the Theoretical Framework

The theoretical framework comprises five aspects: nature of models, multiple models, purpose of models, testing models, and changing models. Whereas the aspects of nature of models and multiple models describe individual ontological and epistemological concepts of models, the aspects of purpose of models, testing models, and changing models refer to cognitive processes while reflecting the act of their use in science (Grosslight et al., 1991; Justi & Gilbert, 2002, 2006).

The aspect of nature of models integrates the aspects of entities and nature (Justi & Gilbert, 2003) as well as the aspect of kinds of models (Grosslight et al., 1991), whereby different views on the nature of science are considered (Hofer & Pintrich, 1997). Here, students compare the model with the original and comment on the extent to which the model is comparable with the original. Three positions are differentiated (Table 1): A model is understood as a replication (level I), as an idealised representation (level II), or as a theoretical reconstruction (level III). Similar aspects were described by Schwarz et al. (2009) as part of metamodeling knowledge and by Oh and Oh (2011) in their review referring to the meanings of a model.

The aspect of multiple models refers to one and the same original being represented by different models (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003). This aspect was deduced from the aspect of multiple models for the same thing (Crawford & Cullin, 2005; Grosslight et al., 1991) and the aspect of uniqueness (Justi & Gilbert, 2003). With the exception of the category of only one model being possible for a phenomenon (Justi & Gilbert, 2003, p. 1375) all described categories were considered. The reason for this exception derived from the fact that studies investigating students' understandings had (so far) provided no clear evidence for this category (Grosslight et al., 1991; Trier & Upmeier zu Belzen, 2009). Regarding the three levels, various explanations can be given for these different models (Table 1): Students justify the existence of several models representing one original by describing differences between the shown model objects such as different materials or

Table 1. The theoretical framework for students' understandings of models and their use in science (Upmeier zu Belzen & Kruger, 2010)

Aspect	Complexity		
	Level I	Level II	Level III
Nature of models	Replication of the original	Idealised representation of the original	Theoretical reconstruction of the original
Multiple models	Differences between different model objects	The original allows the creation of different models	Different hypotheses about the original
Purpose of models	Describing the original	Explaining investigated relationships	Predicting connections between variables
Testing models	Testing the model object itself	Comparing the model with the original	Testing hypotheses about the original with the model
Changing models	Correcting errors in the model object	Revising the model due to new findings about the original	Revising the model due to falsification of hypotheses about the original with the model

dimensions (level I), by arguing that the original allows for building different models in order to present all features of the original (structure and function, level II), or by noticing different hypotheses (level III). A corresponding aspect called multiplicity of scientific models was also noted by Oh and Oh (2011).

In the process of modelling, purpose of models, as well as testing, and changing models need to be reflected upon (Justi & Gilbert, 2002, 2006). Therefore, the process of designing and creating a model is an essential part of all three aspects which is why the aspect of designing and creating models (Crawford & Cullin, 2005; Grosslight et al., 1991) is not distinctly definable. Therefore, this aspect is integrated into the aspects of purpose of models, testing, and changing models. For the aspect of purpose of models, Grosslight et al.'s (1991) and Crawford and Cullin's (2005) aspect of purpose of models as well as the aspects of use and prediction described by Justi and Gilbert (2003) were combined. Three diverse purposes are differentiated: models can serve to show facts of the original (level I), to describe and explain a known relationship in the original (level II), and as an instrument to predict the behaviour of the original (level III, Table 1). Oh and Oh (2011) generated the aspect of purpose of modelling and Schwarz et al. (2009) generated the aspect of using models to explain and to predict what can be assigned to the purpose of models.

The aspect of testing models describes and structures the process of testing models. Here, the category of testing within the aspect of purpose of models described by Grosslight et al. (1991), the aspect of accreditation by Justi and Gilbert (2003), and the aspect of validating/testing by Crawford and Cullin (2005) were conjoined. At level I, the model object itself is tested. At level II, the model is compared with the original and at level III, hypotheses about the original are tested with the model (Table 1).

Changes to the model can be made on the basis of test results. Therefore, models are changeable and temporary by definition (Grosslight et al., 1991; Mahr, 2009; Oh & Oh, 2011). Reasons for changing models are errors in the model object itself (level I), new information about the original (level II), or the falsification of a hypothesis about the original with the model (level III, Table 1). The aspect of changing models is derived from the aspects of changing models (Crawford & Cullin, 2005; Grosslight et al., 1991) and time (Justi & Gilbert, 2003). With regard to other studies, the aspects of testing and changing models have also been elicited by other scholars (Oh & Oh, 2011; Schwarz et al., 2009).

Objective, Research Questions, and Hypotheses

Since the theoretical framework (Upmeier zu Belzen & Krüger, 2010) is a theory-based combination of empirical studies, this study meets the demand of empirically evaluating the structure of this framework. Pursuant to this objective, we aimed to address the following main research question.

To what extent are students' understandings of models and their use in science sufficiently and adequately described by the theoretical framework?

- Which understandings of models and their use in science can be identified and described for each aspect and level of the theoretical framework?
- Which understandings of models and their use in science can be identified in addition to the theoretical framework?
- How frequent are the described understandings of models and their use in science regarding the examined sample?

Since the theoretical framework (Upmeier zu Belzen & Kruger, 2010) brings together empirical studies of Crawford and Cullin (2005), Grosslight et al. (1991), and of Justi and Gilbert (2003), identified categories relating to different understandings of models and their use in science were expected to be sufficiently and adequately described by the theoretical framework. In addition, it was likely that categories at level I and II were more frequently used than categories at level III because international and national studies (Grosslight et al., 1991; Trier & Upmeier zu Belzen, 2009) indicated that students' understandings of models and their use in science differed from a scientific understanding.

Research Design and Method

Instruments

For this study, 15 open-ended test items (three items for each aspect of the theoretical framework) were used to elicit students' understandings of models and their use in science. These 15 open-ended test items result from an item evaluation process involving researchers of biology education and a total of 1,231 German students (seventh to tenth grade students, 12 – 18 years old) in two preceding

studies (Grünkorn & Krüger, 2012; Grünkorn, Upmeier zu Belzen, & Krüger, 2011). Considering the objective of the research study described here (the empirical evaluation of the theoretical framework) test items are required that can be interpreted as indicators for students' understanding of models and their use in science (cf. unified framework for validity, Kane, 2001). To this end, the open-ended items were tested in these two preceding studies for understandability of the items and whether the students' answers represent at least the three levels of the theoretical framework and could be assigned to a certain aspect. Focus was less placed on a selection of certain concrete problem contexts or specific model types (Boulter & Buckley, 2000), but rather on meeting the aforementioned requirements.

The open-ended item format was chosen since the answers can be formulated by the respondents themselves and are not determined by written response options

Table 2. Standardised stimuli for each aspect of the theoretical framework which have been used in the open-ended test items

Aspect	Standardised stimuli	Concrete context used in the items
Nature of models	Describe the extent to which this model of the [original] looks like the [original] Give reasons for your opinion	<ul style="list-style-type: none"> – Theoretical reconstruction of a biomembrane structure – Theoretical reconstruction of a Tyrannosaurus rex – Theoretical reconstruction of a Neanderthal man
Multiple models	Explain why there are multiple models for one [original] Give reasons for your opinion	<ul style="list-style-type: none"> – Different biomembrane structures – Different human gullet structures – Different taste maps of the human tongue
Purpose of models	Describe what purpose this model of the [original] serves	<ul style="list-style-type: none"> – Forest (different plants in a pot of soil) – Ocean (different forms of life in a water bowl) – Gut (pig gut filled with fluid lying in a bowl)
Testing models	Explain in detail how people can test if the model of the [original] serves its purpose	<ul style="list-style-type: none"> – Orientation skills of a beetle – Flying skills of a seed – Flying skills of a dragonfly
Changing models	Name reasons why this model of the [original] could be changed	<ul style="list-style-type: none"> – Flying skills of a dragonfly – Orientation skills of a crab – Food digestion process in the human mouth

Note: Corresponding to the context the term [original] is substituted (cf. Figure 1).

(Rost, 2004). Compared to interviews, which had been conducted by Trier and Upmeier zu Belzen (2009) to analyse students' conceptions of models based on this theoretical framework on a deeper level, open-ended items allow for identifying different understandings of models as well but they relate to a larger sample size. Thus, an overall empirical evaluation of the theoretical framework is possible.

The theoretical framework served as a basis for the item development to the extent that the items were designed for each aspect and that they combined both understandings of models and their use in science: illustration of something in which the model is seen as a medium and the application for something in which the model is used as a method (Gilbert, 1991; Mahr, 2009). The works of Grosslight et al. (1991) as well as an interview study conducted by Trier and Upmeier zu Belzen (2009) served as primary sources for the development of the initial stimuli. In this presented study, the items are situated in different concrete problem contexts (Table 2) since the concept of competence is defined as context-specific for a domain (cf. Klieme et al., 2008). Furthermore, the validity of the context-independent approach by Grosslight et al. (1991) has been discussed in the field of models and their use in science (Krell, Upmeier zu Belzen, Krüger, 2012; Sins, Savelsbergh, van Joolingen,

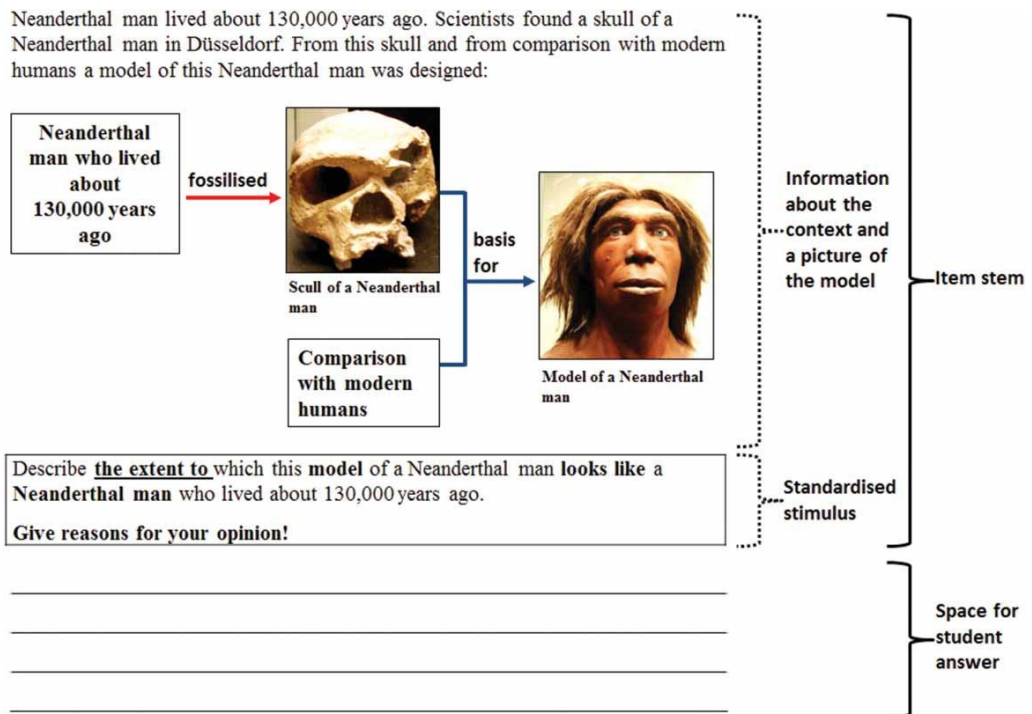


Figure 1. The structure of the open-ended test items exemplified by an item for the aspect of nature of models. The shown item is situated in the context of a theoretical reconstruction of a Neanderthal man (# Pictures: Museum für Naturkunde, Berlin)

& van Hout-Wolters, 2009). Here, the term context relates to situations that are relevant for models in the domain biology (cf. Klieme et al., 2008).

Regarding the structure of the items as exemplified in Figure 1, each item consists of an item stem and the response format—in this case, a blank space for a student’s answer (Rost, 2004). The item stem gives essential information about the context, shows pictures of one or several (concerning the aspect of multiple models) model(s) and ends with a standardised stimulus (Table 2). The stimulus for each aspect only differs according to the context in which the item is situated. Regarding the model type, mostly real objects, technical models (e.g. a flying dragonfly-model), or theoretical reconstructions (e.g. Neanderthal man model; Grünkorn et al., 2011) are used.

Sample and Testing Procedure

The study was conducted with a total of 1,177 seventh to tenth grade students (11–19 years old; 48 different school classes) in Germany (Gymnasium¹, Table 3). The sample of different ages and grades was chosen to identify a wide range of different students’ understandings of models and their use in science. Only thus, it is possible to evaluate whether the theoretical framework reflects learners’ understandings sufficiently. Beyond grade 10, no upper grades were selected for the study because the educational standards in Germany refer to students at the end of grade 10 (KMK, 2005). No further selection criteria were used.

Since one student could not answer all 15 test items and to control order effects, a balanced incomplete block design was developed (Giesbrecht & Gumpertz, 2004). Consequently, the item pool was distributed across 35 test booklets providing each student with three open-ended items and the researchers with about 235 student answers per item (a total of 3,531 student answers). In addition, each booklet included auxiliary variables (such as age, sex, and grade). This block design is appropriate for identifying learners’ understandings of models and their use in science on an aggregate level rather than on an individual level. The objectivity of application is ensured by written instructions in a test manual containing information about the aim of the research project, the item format, standardised answers to frequently asked questions, and the testing procedure.

Table 3. Demographic data about the sample of the study

Grade	n	Sex distribution			Age				
		♀	♂	n. i.	Min.	Max.	M	SD	n. i.
7	313	160	150	3	11	14	12.7	0.57	5
8	293	143	149	1	12	16	13.83	0.62	4
9	283	166	113	4	13	17	14.67	0.62	2
10	288	157	130	1	14	19	15.95	0.69	1

n. i. = no information.

Development and Evaluation of the Category System

The data were analysed by qualitative content analysis according to Mayring (2010), assisted by the qualitative data analysis software MAXQDA (Version 10). This software is commonly used in many academic fields such as sociology, psychology, and educational science and it permits easy scoring, structuring, and restructuring of large amounts of text. Furthermore, it allows for coding a given text passage in any number of ways, highlighting interesting text segments, making notes for reference and ideas directly in the text, and exporting all coding in an excel file. Only the above described functions of the software were used in this study.

As a first step of the qualitative content analysis according to Mayring (2010), the students' answers were manually digitised and colloquial expressions were smoothed while the style of the answers was left unaltered. The changes made within a student answer such as the integration of additional words were marked with square brackets to maintain the integrity of students' responses (cf. Table 4, student Qb417). For the purpose of providing evidence in this article, prominent student answers were translated from German into English by a professional translator (J. P.) and checked by two German researchers who regarded them as accurately translated.

As a second step, similar students' answers given to the respective aspect of the theoretical framework were grouped in categories (called inductive approach, Mayring, 2010). For instance, student Qb64 responds to the standardised stimulus question why there are different models for one biomembrane concerning multiple models with 'Models B and C vary in their solidity. Model B is stiffer while Model C is very flexible'. This student compared the model objects and argued with different characteristics or material properties of the shown models. Another student replied to this question by saying 'Model A is not 3D and models B and C are 3D' (student Qb44). In this case, the student also compared the model objects with each other but argued with different methods of presentation. Those and similar answers were summarised in the category different model object properties (cf. Table 5). Answers that showed content-related incorrectness and/or inadequate planning of experiments were equally analysed and not excluded because the study focused on students' understanding of models and their use in science. The whole category system had already been developed in preceding studies and was revised, discussed, checked, and refined ahead of the study presented in this article.

As a third step, the identified categories were compared with the content and definition of the respective aspect of the theoretical framework (cf. Table 1, Upmeyer zu Belzen & Kruger, 2010). This was done to evaluate whether the students' understandings (categories) of models and their use in science are sufficiently and adequately described by the theoretical framework. If the framework covered the student's understanding, the category was assigned—otherwise it was added to the structure of the framework (called deductive

approach, Mayring, 2010). With regard to the given example of the aspect of multiple models, the category different model object properties (Table 5) focuses only on differences between the model objects and refers to models as a medium (Table 1). Therefore, the described category could be assigned to level I within the aspect of multiple models. The assignments and additions of the categories to the theoretical framework were discussed several times until a consensus was reached with eight researchers of biology education who are experts on the theoretical framework.

As a fourth step, the assignments of students' answers to the categories were also evaluated. For this purpose, three additional trained raters who were familiar with the category system were consulted. To ensure objectivity of the analysis and interpretation, the raters used a coding manual containing information about the items, coding rules for the respective aspect, as well as a detailed description of the category system with prominent student answers from preceding studies. Fifty per cent of the students' answers were double coded and discrepancies in coding were resolved by discussion. Cohen's kappa (k) was used to measure the level of agreement between raters (Fleiss & Cohen, 1973). An agreement between the raters was achieved if all raters assigned the identical text passage within a student answer to the corresponding category. The assignment of an identical text passage to a different category or an additional assignment of text passages to categories by only one rater were noted as discrepancies in coding. The interrater reliability for the assignment of the students' answers to the categories ranged from $k = .81 - .90$ which is characterised as excellent according to Fleiss and Cohen (1973).

Descriptive Analysis

The percentage of each category was calculated based on the number of students who dealt with the respective aspect and the number of coding for each category. The frequencies are presented in Tables 4–10. The sum of frequencies of all categories within each aspect may exceed 100% (cf. Table 8) because the open-ended items allowed for multiple responses in all three levels of the respective aspect. That means one student could present different understandings (categories) in his/her response. As some students did not respond to the open-ended items or their answer did not correspond to the stimulus, the sum of frequencies of all categories within each aspect may not reach 100% (cf. Table 4).

Furthermore, distributions of students across the three levels of each aspect of the theoretical framework are presented to clarify the discussion. For this purpose, all achieved and served levels of a student within each aspect were noted. A level that was mentioned several times by one student was counted only once. The calculation of the percentage of each level was based on the number of students who dealt with the respective aspect and the number of students who showed this level in their response.

Analysis of Student Answers

The descriptive analysis only showed minimal differences between the grades seven to ten, therefore, the different grades were not further considered. Based on 3,531 student answers, 41 categories could be identified in this study describing different students' understandings of models and their use in science.

Categories of Models and their Use in Science and their Frequencies within the Levels of Each Aspect

Twenty-nine out of 41 identified categories in students' responses could be assigned to the respective aspects and the three levels of the theoretical framework. Tables 4–8 summarise the range of categories described for each aspect and their frequencies in the database. The last column of Tables 4–8 presents student answers for each category to provide evidence for each identified category and to substantiate our claims.

Table 4: For the aspect of nature of models, students compared the model with the original and commented on the extent to which the model was comparable with the original. Regarding the frequency distribution across the three levels, the majority of students (69 %) could be assigned to level I. They characterised the model as a copy (e.g. student Qb366, Table 4), as a model with great similarity to the original (e.g. student Qd1073, Table 4), or as a model representing their own (non-) subjective conception of the original (e.g. student Qd915, Table 4). In this context, some students showed a high level of confidence in scientists and scientific work (e.g. student Qa485, Table 4). Furthermore, prior knowledge or personal experience with the original might play an essential role when judging a model. For instance, student Qa317 responded:

Yes, the model looks like a real dinosaur because the dinosaurs back then had also teeth, a head, and a tail like the model shows. I know that because I went to a museum where they had dinosaurs and they look the same.

About 17 % of the students (level II) mentioned that only certain features resembled the original of the model (e.g. student Qa97, Table 4), or understood the model as a possible variant among many (e.g. student Qb417, Table 4), or as a focused representation of something (e.g. student Qb6, Table 4). Only 4 % of the students answered at level III by referring to models as theoretical reconstructions or ideas (e.g. student Qb24, Table 4).

Table 5: The aspect of multiple models related to the question why one original is represented by different models. Approximately half of the students (44 %) compared the model objects describing different model object properties such as different methods of presentation (e.g. student Qc62, Table 5), model features (e.g. student Qb64, Table 5), and construction options (e.g. student Qb311, Table 5). An explicit reference to the original was not made at level I. Almost one-third of the students (31 %, level II) focused on the complexity of the original, and mentioned the necessity of having

Table 4. Described categories, their frequencies (%), and student answers within the three levels of the aspect of nature of models

Name and description of category	Evidence from student answers (student ID no.)	%* (n _{students} = 692)
Level I		
Model as copy	‘The Tyrannosaurus rex looked like the model’ (student Qb366)	27
– Matches the original		
– Enlarged/reduced scale copy of original	‘The model of a biomembrane is an enlarged copy of a biomembrane’ (student Qb502)	
– Accepted as scale model of the original, because there is great confidence in science, in the scientific method, or in the scientists	‘The Neanderthal man looked as it does in the model, because many biologists have certainly worked on this model. These people know what they are talking about, so he looked this way’ (student Qa485)	
Model with great similarity	‘The model of the biomembrane is very similar to the real biomembrane. Both have a surface layer that holds it all together and both have tissue in the centre’ (student Qd1073)	36
– Resembles the original		
– Nearly scale model of the original due to dissatisfaction with the modelling process	‘The model resembles the Neanderthal man. Only the place on the model where the eyebrows are has to be pushed forward a little, because that’s how it is in the skeletal findings. That has to happen; otherwise, the model would be incorrect’ (student Qd567)	
Model represents a (non-) subjective conception of the original	‘I don’t think the model is correct. The real Neanderthal man looked more like an ape. That’s how I imagine one’ (student Qd915)	6
– Compares and judges the model based on prior knowledge of, personal experience of, or subjective conceptions about the original		
Level II		
Parts of the model are a copy	‘The skeletal findings and the model have the same head shape. It is unknown whether the hair back then was the way it is now. Nothing can be stated with certainty about the eyes, either. On the whole, one can only comment on the shape. Colour and such remains “unknown”’ (student Qa97)	9
– Only certain features resemble the original; other features cannot be judged due to paucity of information or knowledge about the original		

(Continued)

Table 4. Continued

Name and description of category	Evidence from student answers (student ID no.)	%* (n _{students} = 692)
Model as a possible variant – Might resemble the original (or not); abstract statements about similar properties – One conceivable version among many, but less well founded	‘The model is comparable in terms of its shape. Nonetheless, one cannot assume that the Neanderthal man really looked like this’ (student Qa164) ‘Yes, it [the biomembrane] might look like that, but it might also look like this [picture drawn by student]’ (student Qb417)	7
Model as focused representation – Focused on one element of the original, highlights certain traits/properties	‘The model only shows the essential parts of a real biomembrane. The main traits, structures and colours are shown here’ (student Qb6)	0.3
Level III Model as hypothetical representation – Presents a justified hypothesis about the original, possible similarity between original and model is discussed	‘No one can know for certain what a living Tyrannosaurus rex looked like back then. Scientists can only make assumptions about how it looked. They analyse the skeleton and use that to calculate how its body might have been constructed’ (student Qb24)	4

*Each frequency is based on all noted students’ understandings and is calculated with reference to the number of students who dealt with this aspect.

different models in order to show every aspect of the original as exemplified by student Qb50: ‘... Model A shows the structure of the biomembrane precisely and Model C shows it roughly. Model B shows the idealised function’. Here, an explicit reference to the original was made. For students it was rather rare (8%, level III) to justify different models of one and the same original with various assumptions or ideas about the original (e.g. student Qb5, Table 5). The scientific purposes of these different hypotheses about the original such as starting points for discussions, comparison of different assumptions, and testing assumptions with the models were named by only a few students (e.g. student Qd353, Table 5).

Table 6: In response to the purpose of models, a good proportion of the students gave answers at level I (52%) and II (50%). At level I, the purpose of models was seen as showing facts without using the model to recognise a certain relationship between different aspects in the original (e.g. student Qa403, Table 6). The latter was characteristic for level II in which the model served to identify (e.g. student Qa217, Table 6) or explain relationships (e.g. student Qb175, Table 6) in order to understand known facts. With regard to level III, 24% of the students stated that models were instruments to examine ideas by testing hypotheses about the original. Here, the facet of gaining new information and learning something new about a specific phenomenon

Table 5. Described categories, their frequencies (%), and student answers within the three levels of the aspect of multiple models

Name and description of category	Evidence from student answers (student ID no.)	%* (n _{students} = 705)
Level I		
Different model object properties	‘The tongue can be shown a little differently. No colours are used in Model C, for example, whereas they are in Models A and B’ (student Qc62) ‘Models B and C vary in their solidity. Model B is stiffer while Model C is very flexible’ (student Qb64) ‘The different elements are easy to recognise in Model A. The rods on Model C are thinner than in B’ (student Qb311)	44
– Differing methods of presentation (2D or 3D, different colours, etc.)		
– Differing model features (moveable or immovable, soft or hard, large or small, etc.)		
– Differing construction options (thin or thick materials, separated elements or one piece, etc.)		
Level II		
Focus on different aspects	‘Since each of these models highlights something different, there are different models. Model A focuses on the different elements and the structure, while Model B and C look more at the construction of a biomembrane’ (student Qd744)	31
– The complexity of the original allows diverse perspectives or ways of focusing on the original (interior or exterior, profile or cross-section, structure or function, diverse sections or states of the original, etc.)		
Level III		
Different assumptions	‘Since there are various theories/ideas about the human oesophagus, there will also be alternative models. Scientists might have other opinions’ (student Qb5) ‘The persons have drawn different conclusions from their observations, which is why there are different models of this biomembrane’ (student Qd468)	8
– There can be various assumptions and ideas about the original; different models are valid at the same time		
– Differing interpretations of the data		
Different assumptions with prospects of application	‘Probably because there are many people who have researched this and so different ideas developed which still have to be examined in studies’ (student Qd353)	1
– Differing assumptions about the original are named after scientific purposes (basis of discussion, comparison of different assumptions, testing assumptions with the models, etc.)		

*Each frequency is based on all noted students’ understandings and is calculated with reference to the number of students who dealt with this aspect.

Table 6. Described categories, their frequencies (%), and student answers within the three levels of the aspect of purpose of models

Name and description of category	Evidence from student answers (student ID no.)	%* ($n_{\text{students}} = 706$)
Level I		
Model for showing the facts – Presenting the facts	‘The model shows the different plants that grow in a forest’ (student Qa403)	52
Level II		
Model to identify relationships – Describing relationships between different aspects in the original and serving to understand known facts	‘The model shows that it is possible to observe how the leaves and blossoms develop and spread’ (student Qa217)	45
Model to explain relationships – Describing and explaining relationships between different aspects in the original and serving to understand known facts	‘It is meant to demonstrate that the sea is a good habitat for animals e.g. fish and plants. It also shows and explains how the sea constitutes a “circulatory chain”. The plants could not survive without the oxygen and the water; the fish could not live without the plants’ (student Qb175)	6
Level III		
Model to examine abstract ideas – Serving as an instrument to test hypotheses about the original; general ideas are mentioned	‘Perhaps so that certain tests can be viewed and carried out so that the effects of certain differences such as temperature can be studied’ (student Qd1168)	19
Model to examine concrete ideas – Both testing hypotheses about the original and serving to draw conclusions about the original; concrete ideas are mentioned – Serving to transfer findings about the original to other phenomena	‘One could also test which plants grow best and most quickly in which types of soil and compare these results. It might be that a certain plant draws so many nutrients out of the soil that there are less available for another. If the model were to prove this, we would know that these types of plant should not be planted too closely together’ (student Qd263) ‘Such a model is probably there to test whether plants can “multiply”. This information helps to say something about the forest and to capture this situation in figures or charts. Then we could relate this to other ecosystems, like the sea and such’ (student Qb291)	5

*Each frequency is based on all noted students’ understandings and is calculated with reference to the number of students who dealt with this aspect.

Table 7. Described categories, their frequencies (%), and student answers within the three levels of the aspect of testing models

Name and description of category	Evidence from student answers (student ID no.)	%* ($n_{\text{students}} = 711$)
Level I		
Testing of material – Testing the resistance of the material (flexibility, stability, elasticity, weight, etc.)	‘One should test if the material of the model is strong enough to remain unharmed by something such as wind’ (student Qd793)	6
Testing of basic requirements – Naming fundamental requirements for that model	‘For starters, the model should be able to fly in any case. Otherwise, I don’t think the model would be very good’ (student Qa4)	28
Level II		
Comparison between original and model – Comparing the properties (structure and/or function) of the original with those of the model	‘The model has to be compared to a real beetle’ (student Qc69)	33
Comparison and matching of original and model – Both comparing properties and describing the necessary adjustments for congruity between the model and the original; naming criteria for a good model	‘The model can be tested for its dimensions, its weight. The structure of the model must match the original or it isn’t suitable’ (student Qb206)	34
Level III		
Testing hypotheses – Testing hypotheses about the original using the model and listing general ideas for studies	‘This model could simulate the flight of such a seed. Such simulations would show where the seed flies to and how it gets implanted into the soil. The model could also be used to test the effects the impact has on the soil, on the flight, and on the seed’ (student Qb278)	6
Testing of hypotheses with research designs – Describing a concrete application for the model (research design) to test a hypothesis about the original	‘One has to try to obtain videos of the original flight manoeuvres and attempt to recreate and compare these with the model in a wind tunnel to see if the model behaves as the original. If so, one has to change the environmental influences in the wind tunnel to determine what the dragonfly needs to fly’ (student Qb200)	2

*Each frequency is based on all noted students’ understandings and is calculated with reference to the number of students who dealt with this aspect.

Table 8. Described categories, their frequencies (%), and student answers within the three levels of the aspect of changing models

Name and description of category	Evidence from student answers (student ID no.)	%* (n _{students} = 712)
Level I		
Alterations to improve the model object	‘The only reason why most models are changed is because their movement and functionality can be improved’ (student Qa378) ‘To change the model of the dragonfly, a new technology is needed that allows the model to stay up without needing to attach a booster for uplift on the long back legs’ (student Qb521)	22
– Optimising the functioning/aesthetics of the model object – Optimising the technology of model creation		
Alterations when there are errors in the model object	‘I think it’s because errors are always being found which need to be corrected’ (student Qd629) ‘Perhaps the wings have to be made out of harder materials; otherwise they cannot resist the pressure during flight’ (student Qb508)	5
– Fundamental considerations for fixing errors in the model – Referencing concrete, incorrect properties of the model (e.g. defective materials)		
Alterations when basic requirements are not met	‘If the model is meant to fly and it doesn’t, then the scientists have to work on it’ (student Qb496)	9
– Reviewing the basic requirements of each model and correcting defects if necessary		
Level II		
Alterations when model does not match the original	‘The model doesn’t look exactly like a crab. The legs of a real crab are longer. The body of a crab is somewhat narrower. This is not the shape of a crab. That should definitely be changed, because it has to match the real crab’ (student Qd1010)	61
– Optimising how the (structure and/or function of the) model matches the original with consideration of the necessary congruity between the original and the model		
Alterations due to new findings about the original	‘In a few years, we will have better technology, so we can learn more about the dragonfly. The model could be changed when something new about the dragonfly is discovered’ (student Qd1199)	4
– Integrating new findings about the original into the model; improved technology leads to new findings about the original		
Alterations due to changes in the original	‘There are always changes in biology and in history. The same is true of the crab. Evolution changes the environment and animals have to adapt again. Changes to the environment force animals to change as well. That’s why the model can be changed’ (student Qd1165)	5
– Reflecting changes (e.g. individual developments) or advancements (e.g. evolutionary adaptation) in the original as new information in the model		

(Continued)

Table 8. Continued

Name and description of category	Evidence from student answers (student ID no.)	%* (n_{students} = 712)
Level III Alterations due to findings from model experiments – Adjusting the model to reflect findings about the original based on a model experiment or falsification of the hypothesis behind the model	‘If tests of a flying object show that the model flies completely differently than thought or than a real dragonfly does, then something could be changed on the gliding surfaces. The scientists may have had a different assumption’ (student Qd352)	1

*Each frequency is based on all noted students’ understandings and is calculated with reference to the number of students who dealt with this aspect.

played a decisive role, as exemplified by student Qb291 (Table 6). At level III, students noticeably more often expressed general ideas (category model to examine abstract ideas, 19 %, Table 6) than concrete ideas (category model to examine concrete ideas, 5 %, Table 6).

Table 7: For the aspect of testing models, students explained how people could test whether the model served its purpose. The most popular responses (level II, 68 %) to this question were to test a model by comparing the original to the model (e.g. student Qc69, Table 7) or by comparing properties and describing necessary adjustments for congruity between the original and the model (e.g. student Qb206, Table 7). Considering the frequencies within the different levels, level II (68 %) was followed by level I (30 %). At level I, students tested mainly the material of the model object for robustness (e.g. student Qd793, Table 7) or tested if the model objects fulfilled basic requirements (e.g. student Qa4, Table 7). A few students’ responses assigned to level III (8 %) concerned testing hypotheses about the original by describing general (e.g. student Qb278, Table 7) or concrete (e.g. student Qb200, Table 7) ideas for studies with the aim of gaining purposeful new insights into biological topics. Similar to the aspect of purpose of models (Table 7), students mentioned general ideas (category testing hypotheses, 6 %, Table 7) more often than concrete ideas (category testing hypotheses with research designs, 2%, Table 7).

Table 8: The aspect of changing models pursued the question of what conditions could lead to changing a model. Similar to testing models, most students (68 %) described understandings that could be assigned to level II of the theoretical framework paying attention to the original: a majority of students (61 %, category alteration when model does not match the original, Table 8) at level II argued that a model is changed when it does not match the original in terms of structure and/or function (e.g. student Qd1010, Table 8). Only a few students named reasons such as alteration due to new findings about the original (4 %, e.g. student Qd1199, Table 8) or alteration due to changes in the original (5 %, e.g. student Qd1165, Table 8). Approximately one-third of the students (31%, level I) referred to alterations that improve (e.g. student Qa378, Table 8) or fix the model object when defects were found in

the model object itself (e.g. student Qb508, Table 8) or the basic requirements were not met (e.g. student Qb496, Table 8). Students at this level did not mention the original as a reason for a change; they rather focused on general technical issues in the model object, as exemplified by student Qb521: ‘The model of the dragonfly is changed when a new technology allows the model to stay up without needing to attach a booster for uplift on the long back legs’. Rarely (1 %) did the students change a model due to findings from model experiments in which their hypothesis, implicitly stated by students in their answers, was rejected (level III, e.g. student Qd352, Table 8).

Additional Categories of Models and their Use in Science to Each Aspect

For the aspect of nature of models, no additional categories could be identified within the data. However, concerning the aspects of multiple models, purpose of models, testing models, and changing models, additional categories could be generated (Tables 9 and 10).

Table 9: The aspect of multiple models and its three levels refer to the understanding of one and the same original being represented by different models. Therefore, students were confronted with three different models of one original in the open-ended items. Special attention was paid in the development of the open-ended items to ensure that distinct references to one original were made in the item stems and obviously different models were shown (such as different foci, colours, material, and dimensions). Still, 14 % of the students gave one of the following responses when asked about the presence of multiple models (Table 9): (a) all models were the same (e.g. student Qd240, Table 9), (b) various models representing different originals were shown (e.g. student Qb331, Table 9), or (c) only one model was the final and correct one (e.g. student Qd643, Table 9). Those students rejected the existence of multiple models as representations of one original and conceived only one model as representation of an original. By giving one of the listed responses (Table 9), they stuck to their understanding. As the aspect of multiple models demands the acceptance and understanding of multiple models referring to one original, all three facets can be seen as an initial understanding of this aspect.

For the aspect of testing models, 1% of the students (e.g. student Qc71, Table 9) thought it was unnecessary to test a model and expressed their perplexity about why and how a model is tested. Students kept this understanding although the purpose of the presented models could plausibly be used for testing. The concept of testing models includes the acceptance and necessity for testing, which is why this kind of response can be interpreted as an initial understanding and level of this aspect.

Similarly, an initial understanding and level could be identified for the aspect of changing models reflecting the opinion that models should not be changed. Students (10 %) who showed this understanding responded by either rejecting changes of a model (e.g. student Qd341, Table 9), or by changing the presented model to represent another original (e.g. student Qd954, Table 9). Since the concept of this aspect

Table 9. Initial understandings of the aspects of multiple models, testing models, and changing models exemplified by student answers and their frequencies (%)

Name and description of category	Evidence from student answers (student ID no.)	%*
Multiple models ($n_{\text{students}} = 705$)		
All models are the same – All models are or show the same; no description of differences between models	‘All three models show basically the same. I don’t know why there should be different models at all. That makes no sense’ (student Qd240)	4
Various models of different originals – Each model represents a different original	‘One might also make three different models to show the biomembranes of different life forms, e.g. a human being, a bird and a cow’ (student Qb331)	13
Only one final and correct model – Only one of the various models is final and correct; the others are incorrect – Only one model is the final model; they are not valid contemporaneously	‘Perhaps there is only model which is the best model. All the others are wrong. There is just one correct model. How else should this work?’ (student Qd643) ‘I think that two models are old models. At that time, one did not have all information like we have today. However, one model is the final model which happens to be true. I’ve seen it in my textbook’ (student Qd321)	1.4
Testing models ($n_{\text{students}} = 711$)		
No testing of models – Rejecting model testing in general or of this model	‘Why should this beetle model even be tested? I don’t think it’s necessary’ (student Qc71)	1
Changing models ($n_{\text{students}} = 712$)		
No reason for alterations – Rejecting changes to a model	‘I don’t think the model should be changed’ (student Qd341)	3
Alteration of how different originals are represented – Creating different models for different originals; each original is represented by its own model	‘Because not all dragonflies are alike and models can be made for different dragonflies’ (student Qd954)	7

*Each frequency is based on all noted students’ understandings and is calculated with reference to the number of students who dealt with this aspect.

demands the acceptance of changing a model and that the creation of a model as well as the process of changing a model are always linked to a specific purpose and a specific original, these categories were not assigned to the three levels of the aspect of changing models.

Table 10: For the aspects of purpose of models and changing models, further categories could be generated that were not considered as initial understandings of these aspects. As the question regarding purpose of models was formulated in a rather open way

Table 10. Further categories, their frequencies (%), and student answers for the aspect of purpose of models presenting everyday understandings, and a category of the aspect of changing models showing an alternative understanding

Name and description of category	Evidence from student answers (student ID no.)	%*
Purpose of models ($n_{\text{students}} = 706$)		
Model as toy – Serving for pleasure, interest, and/or enjoyment	‘There might be children who are interested in it and will have fun with it, so this model was developed’ (student Qd580)	0.4
Model for decorative purposes – Serving to beautify a room or building – Contributing to the well-being of people and other life forms	‘It might be that the forest model can be put to 2 good purpose decorating the home’ (student Qd1137) ‘Some people simply find it [the model] relaxing when they aren’t on holiday or simply want to daydream’ (student Qd1035)	2
Model as replacement object – Serving as prosthesis or replacement organ	‘Its purpose might be that it was developed as 1 a replacement organ. If someone has a defective intestine, it could be replaced with an artificially manufactured, fully functional one’ (student Qd31)	1
Model as blueprint – Serving as a small-scale blueprint for presenting the future object	‘This model is meant to be used like a map. 0.1 One could say that a hotel miniature is to be placed on the model and then one could build the real hotel exactly on the spot represented in the model. It serves for orientation, so to speak’ (student Qd144)	0.1
Model for advertising purposes – Serving as advertisement to reach political, social, and/or environmental goals	‘They want to use it to make the world aware 1 that forests need more water’ (student Qd613)	1
Changing models ($n_{\text{students}} = 712$)		
Alterations of the model conditions – Changes to the model to reflect changes in how it is used or handled (movements of the model, etc.)	‘The model of a mouth could change during 2 rotation’ (student Qd1208)	2

*Each frequency is based on all noted students’ understandings and is calculated with reference to the number of students who dealt with this aspect.

(Table 2), the student responses touched a wide range of purposes. For example, students thought of models as toys (e.g. student Qd580, Table 10), as something for decorative purposes (e.g. students Qd1137 and Qd1035, Table 10), as replacement objects (e.g. student Qd31, Table 10), as blueprints (e.g. student Qd144, Table 10), or as something for advertising purposes (e.g. student Qd613, Table 10). With regard to changing models, a few students had an alternative understanding of

the term change by thinking that a change referred to the alteration of conditions underlying the model (e.g. student Qd1208, Table 10).

Discussion and Conclusion

The purpose of this study was to empirically evaluate the theoretical framework. Before contrasting the findings with the results of other scholars in this field, methodological constraints need to be discussed: the preceding studies have already shown that the 15 open-ended test items can be interpreted as indicators for the understanding of models and their use in science (Grünkorn & Krüger, 2012; Grünkorn et al., 2011). Therefore, the items are suitable for evaluating the structure of the theoretical framework concerning learners' (grades 7–10; 11–19 years old) understandings of models and their use in science. Concerning the target group (grades 7–10) for whom these instruments are designed, the results only allow for reliable statements concerning these particular learners.

For the presented study, a context-dependent approach was chosen since the validity of a context-independent approach by Grosslight et al. (1991) as well as Treagust et al. (2002) have been discussed and questioned in the field of models (Krell et al., 2012; Sins et al., 2009). All items were situated in different concrete contexts (Table 2). Thus, different results concerning the frequencies may be achieved when choosing a context-independent approach or when choosing other contexts (Krell et al., 2012). Since the focus of this study was rather to evaluate the theoretical framework on an aggregate level than to diagnose students' understandings of models and their use in science on an individual level, this issue can be neglected (cf. Leutner, Fleischer, Spoden, & Wirth, 2007).

Bearing in mind that the items are situated in biological contexts, findings cannot—without further investigations—be generalised and therefore only hypothetically be transferred to other science domains such as physics or chemistry. This precaution is based on the fact that we define learners' understandings as competencies. Competencies are acquired and applied in a specific domain (cf. Fleischer et al., 2013; Klieme et al., 2008). Furthermore, in scientific subjects such as biology, chemistry, and physics models are used in different ways. Whereas in biology structural, functional, or dynamic models are often used, abstract and mathematical models predominate in chemistry and physics (cf. Beerenwinkel & Parchmann, 2008; Mikelskis-Seifert & Fischler, 2003).

Empirical Evaluation of the Theoretical Framework

With regard to the main research question whether the theoretical framework (Upmeier zu Belzen & Krüger, 2010) accurately reflects the different aspects and levels of learners' understandings concerning models and their use in science, the hypothesis was only partly confirmed: the students' understandings of the aspect of nature of models and purpose of models are sufficiently and adequately described by the theoretical framework. However, the aspects of multiple models, testing models, and changing models need to

be discussed and, therefore, revised since initial understandings for these aspects were identified. Table 11 summarises the findings of the study and provides a revised theoretical framework, that contributes to research in this field.

Table 11. Revised framework for students' understandings of models and their use in science including levels of complexity and their categories

Aspect	Complexity			
	Initial level	Level I	Level II	Level III
Nature of models	—	Model as copy Model with great similarity Model represents a (non-) subjective conception of the original	Parts of the model are a copy Model as a possible variant Model as focused representation	Model as hypothetical representation
Multiple models	All models are the same Various models of different originals Only one final and correct model	Different model object properties	Focus on different aspects	Different assumptions Different assumptions with prospects of application
Purpose of models	—	Model for showing the facts	Model to identify relationships Model to explain relationships	Model to examine abstract ideas Model to examine concrete ideas
Testing models	No testing of models	Testing of material Testing of basic requirements	Comparison between original and model Comparison and matching of original and model	Testing hypotheses Testing of hypotheses with research designs
Changing models	No reason for alteration Alteration of how different originals are represented	Alterations to improve the model object Alterations when there are errors in the model object Alterations when basic requirements are not met	Alterations when model does not match the original Alterations due to new findings about the original Alterations due to changes in the original	Alterations due to findings from model experiments

The initial understanding of the aspect of multiple models could be described by the conception only one model of an original. Grosslight et al. (1991) indicated that some students did not believe that it was possible to have multiple models representing one original. Justi and Gilbert (2003), however, explicitly described the conception uniqueness (p. 1375) in a study with teachers. The students' responses concerning the uniqueness of models named in this article (categories: all models are the same and only one final and correct model) were similar to findings reported by Justi and Gilbert (2003). Students giving the latter explanation probably understood models as final form devices and evaluated multiple models for correctness (Crawford & Cullin, 2004). Those students' understandings differ from scientific understandings which '... do not ask whether it [the model] is right or not' (Crawford & Cullin, 2004, p. 1382). Contrary to Crawford and Cullin (2004), Grosslight et al. (1991), and Justi and Gilbert (2003), the category of various models of different originals was elicited among students' responses and played an important role when explaining the presence of multiple models representing one original. These results demonstrate that this understanding is also prominent among students in different variations and therefore needs to be considered.

Based on the analysis of the students' responses, a new initial understanding could be identified for the aspect of testing models. Thus, a few students rejected the necessity of testing a model and, therefore, showed no clear awareness of the work of scientists which is predominated by the creation as well as the testing of models (Giere, 2004; Odenbaugh, 2005; Oh & Oh, 2011). The studies by Crawford and Cullin (2004, 2005), Grosslight et al. (1991), and Justi and Gilbert (2003) did not give any indication as to this category. This newly described category is a result from this study and it needs to be considered and focused on in future research.

Regarding the aspect of changing models, some students mentioned the understanding of models being unchangeable (Table 9). This understanding has previously been reported by other studies (Crawford & Cullin, 2004, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Treagust et al., 2002), and it confirms the idea that some students think of models as final form devices (Crawford & Cullin, 2004). Besides clearly rejecting changes to the model, several students explained that a model could be changed in order to present another original. Similar to the category of various models of different originals in the aspect of multiple models, students again referred to different originals instead of focusing on one original. The category also emerges from this study and has not been described by other studies (Grosslight et al., 1991; Justi & Gilbert, 2003).

Based on these results, for the aspects of multiple models, testing, and changing models a fourth level of students' understandings of models and their use in science needs to be added (Table 11). Therefore, the results rather support the approach by Crawford and Cullin (2005) who defined three to four levels for each aspect than the approach of Grosslight et al. (1991), of Justi & Gilbert (2003), or of Upmeier zu Belzen and Krüger (2010) who differentiate between three levels.

Besides describing initial understandings, newly described categories within the three levels of the aspects of purpose of models and testing models could be identified

(Table 11). Unlike the categorisation for the aspect of purpose of models shown in the studies by Crawford and Cullin (2005), Grosslight et al. (1991), and Justi and Gilbert (2003), the categories of models to examine abstract ideas (19 %) and models to examine concrete ideas (5 %) could be described and distinguished within the data (Table 6, level III). Similar categories (Table 7, level III) were described in the aspect of testing models in which hypotheses about the original are tested by either describing general (6%) or concrete ideas (2%). This analysis allowed for the conclusion that students have an abstract and/or concrete understanding and that it is probably more difficult and complex to establish a concrete understanding than an abstract one. This argument is supported by the fact that a qualitative study using thinking aloud protocols to elicit students' understandings (Terzer, 2013) showed that students did have an abstract understanding of models but could not use it to solve the problem in concrete situations with models. Therefore, future research needs to focus on these different understandings.

In summary, the purpose of this study was to evaluate a theoretical framework that brings together different approaches (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003) and is empirically evaluated with a large sample and students of different ages. This overall empirical evaluation did not only provide evidence for students having an initial understanding concerning the 'uniqueness' (Justi & Gilbert, 2003) of models, but newly described categories were identified that need to be considered in future research (Table 11). This conjoint and empirically tested framework can now serve as a basis for future investigations concerning students' understandings of models and their use in science.

Further Categories within the Data

The analysis of students' responses to the aspect of changing models showed that a few students had an alternative understanding of the term changing in this context. They perceived a change as an alteration of the model condition (Table 10). This understanding does not agree with the understanding that is fundamental in the theoretical framework (Upmeier zu Belzen & Krüger, 2010). Although this category is not being considered in the theoretical framework, there is a need to gain access to this understanding by asking students in a separate questionnaire to prevent misunderstandings.

Frequencies of the Generated Categories

We were also interested in how often the described understandings of models and their use in science could be determined in the examined sample on an aggregate level. Regarding the percentage distribution within each aspect of the theoretical framework (Upmeier zu Belzen & Krüger, 2010), students responded more frequently at level I and II than at level III. These findings were as expected and corresponded largely to results of other studies (Grosslight et al., 1991; Schwarz & White, 2005; Trier & Upmeier zu Belzen, 2009). Differences in the percentage distribution were noted in the aspects of multiple models and changing models.

In contrast to the studies of Grosslight et al. (1991), in which students often explained the presence of multiple models with different foci (level II), students in the presented study rather focused on differences between the model objects in terms of colour, shape, dimension, and material (level I). Concerning changing models, a popular response did not concern alterations due to new findings about the original (4 %) but alterations when the model does not match the original (59 %). The comparison between the original and the model played an important role and showed that students were firmly anchored in a medial perspective on their use in science (Oh & Oh, 2011). The differences concerning the percentage distribution for the aspects of multiple models and changing models might have been caused by the fact that this presented study was conducted with a larger sample than the previous studies (Grosslight et al., 1991). Another possible explanation might be that in the presented study, the items were situated in different concrete problem contexts. Sins et al. (2009) noted and a study using forced-choice items (Krell et al., 2012) showed empirically that the context-dependent and the context-independent approaches lead to different results. However, further investigations concerning the impact of different concrete problem contexts and model types on students' performances need to be conducted.

Students in this sample see models less as a method of science (Mahr, 2009; Oh & Oh, 2011) and more from a medial perspective (Mahr, 2009; Oh & Oh, 2011). Possible reasons might be a more frequent use of models as a substitute for the original or as a medium for transmitting information in biology lessons (Crawford & Cullin, 2004; Van Driel & Verloop, 1999). This and the role of teachers in 'distinguishing the positive and negative analogies as clearly as possible' (Hardwicke, 1995, p. 64) might lead to the fact that students primarily focus on the relationship between the model and the original. Students draw comparison to achieve the best possible match without considering the purpose of the model.

Final Implications

With regard to science education research the following contributions and implications can be derived from the results of this study: first, to assess and diagnose the current state or development of students' understandings of models and their use in science, more sensitive instruments might be helpful. For this purpose, the comprehensive category system can provide student-based understandings and could serve as a basis for the development of those instruments. Second, a conjoint and empirically tested framework for students' understandings of models and their use in science is now available. This framework is applicable to future research in this field, for instance, by having created a precondition to address the demand raised by Louca and Zacharia (2012). In their paper, they requested that research needs to focus on investigating the relationships between students' understandings of models, their practical skills, and other components such as the social facet that influence students' processes of modelling. This research gap can be addressed by developing frameworks for each component

that allow for assessing these components separately and consequently for investigating the relationships between them. In terms of providing an empirically tested framework for assessing students' understandings, this was accomplished by this study. Regarding the other facets, conjoint and consensual frameworks need to be developed. For this purpose, the works of Louca et al. (2011) as well as Schwarz et al. (2009, 2012) already provide valuable information on students' practical skills. Third, the conjoint and empirically tested framework and the comprehensive category system presented in this article can be used to assess and investigate the cognitive facet of models, i.e. students' competencies of models and their use in science. As the concept of competencies is defined as a basis for successfully solving problems in various situations (cf. Klieme et al., 2008; Weinert, 2001), the students' achieved level gives valuable information to teachers and researchers concerning the extent to which they are able to solve problems with models in certain modelling situations.

Concerning science education, several studies already dealt with learning progressions for scientific modelling and showed that constructing models and reflecting about models help to facilitate students' understandings of models and modelling (Baek, Schwarz, Chen, Hokayem, & Zhan, 2011; Schwarz et al., 2009). This study contributes to the field by giving further implications regarding students' understandings of models and their use in science: first, we suggest being aware not only of the initial students' understandings of models and their use in science but also of the other understandings. Thus, the category system might be used in teacher training to assist the teaching staff to keep abreast of their students' different understandings of models and their use in science (cf. Fleige, Seegers, Upmeier zu Belzen, & Krüger, 2012). Second, in this context, a reflective use of historical models in school needs to be discussed as already demanded by Justi and Gilbert (2002). The findings of this study suggest that epistemological processes in science and the hypothetical character of models could lead to an elaborated understanding of models and their use in science (Justi & Gilbert, 2003) as exemplified by student Qd468 '... Models could represent different ideas of scientist which changed over time'. However, an unreflective use of historical as well as current models might contribute to the understanding why the presence of multiple models representing one original is rejected. According to student Qd321 (Table 9), the historical processes lead to a final and correct model which does not change. Therefore, it is important to carefully reflect current models and to consider and discuss the conclusions students draw from historical processes. Third, as giving feedback is an integral part of fostering students adequately, the category system might be used by teachers as an instrument to provide students with detailed feedback.

Note

1. Gymnasium is a type of secondary school in Germany in which students finish school after grade 12 or 13 with A-levels (Abitur). The Abitur qualifies them for admission to university.

References

- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy: Project 2061*. New York, NY: Oxford University Press.
- Baek, H., Schwarz, C., Chen, J., Hokayem, H., & Zhan, L. (2011). Engaging elementary students in scientific modeling: The MoDeLS fifth-grade approach and findings. In M. Khine & I. Saleh (Eds.), *Models and modeling* (pp. 195–220). Dordrecht: Springer.
- Beerenwinkel, A., & Parchmann, I. (2008). Metadiskussion über Modelle. Historische Aspekte als Impuls [Meta-discussion about models. Historical aspects as impulses]. *Praxis der Naturwissenschaften Chemie* [Practice of science chemistry], 4(57), 13–16.
- Boulter, C., & Buckley, B. (2000). Constructing a typology of models for science education. In J. Gilbert & C. Boulter (Eds.), *Developing models in science education* (pp. 41 – 57). Dordrecht: Kluwer Academic.
- Crawford, B. A., & Cullin, M. J. (2004). Supporting prospective teachers' conceptions of modelling in science. *International Journal of Science Education*, 26(11), 1379 – 1401. doi:10.1080/09500690410001673775
- Crawford, B., & Cullin, M. (2005). Dynamic assessments of pre-service teachers' knowledge of models and modelling. In K. Boersma, H. Eijkelhof, M. Goedhart, & O. Jong (Eds.), *Research and the quality of science education* (pp. 309–323). Dordrecht: Springer.
- Fleige, J., Seegers, A., Upmeyer zu Belzen, A., & Krüger, D. (2012). Förderung von Modellkompetenz im Biologieunterricht. [Fostering model competence in biology education]. *Der mathematische und naturwissenschaftliche Unterricht* [Mathematical and science teaching], 65(1), 19–28.
- Fleischer, J., Koeppen, K., Kenk, M., Klieme, E., & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms [Modelling of competencies: Structure, concepts and research approaches of the DFG priority program]. *Zeitschrift für Erziehungswissenschaften* [Journal for educational science], 16(1), 5 – 22. doi:10.1007/s11618-013-0379-z
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613 – 619. doi:10.1177/001316447303300309
- Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, 71, 742–752. doi:10.1086/425063
- Giesbrecht, F., & Gumpertz, M. (2004). *Planning, construction, and statistical analysis of comparative experiments*. Hoboken, NJ: John Wiley & Sons.
- Gilbert, S. W. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, 28(1), 73–78. doi:10.1002/tea.3660280107
- Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28(9), 799–822. doi:10.1002/tea.3660280907
- Grünkorn, J., & Krüger, D. (2012). Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells zur Modellkompetenz [Development and test of open-ended test items to empirically assess a model for model competency]. In U. Harms & F. X. Bogner (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik* [Teaching/Learning research in biology education] (pp. 9 – 27). Innsbruck: Studienverlag.
- Grünkorn, J., Upmeyer zu Belzen, A., & Krüger, D. (2011). Design and test of open-ended tasks to evaluate a theoretical structure of model competence. In A. Yarden & G. Carvalho (Eds.), *Authenticity in biology education* (pp. 53–65). Braga: CIEC, Universidade do Minho.
- Hardwicke, A. J. (1995). Using molecular models to teach chemistry: Part 1 using models. *School Science Review*, 77(278), 59–64.

- Hartig, J. (2008). Kompetenzen als Ergebnisse von Bildungsprozessen [Competencies as results of educational processes]. In N. Jude, J. Hartig, & E. Klieme (Eds.), *Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden [Assessing competencies in pedagogical contexts. Theories, concepts, and methods]* (pp. 17 – 25). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67, 88 – 140. doi:10.3102/00346543067001088
- Justi, R. S., & Gilbert, J. K. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. *International Journal of Science Education*, 24(4), 369 – 387. doi:10.1080/09500690110110142
- Justi, R. S., & Gilbert, J. K. (2003). Teachers' views on the nature of models. *International Journal of Science Education*, 25(11), 1369–1386. doi:10.1080/0950069032000070324
- Justi, R. S., & Gilbert, J. K. (2006). The role of analog models in the understanding of the nature of models in chemistry. In P. J. Aebischer, A. G. Harrison, & S. M. Ritchie (Eds.), *Metaphor and analogy in science education* (pp. 119–130). Dordrecht: Springer.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38, 319 – 342. doi:10.1111/j.1745-3984.2001.tb01130.x
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., . . . Vollmer, H. J. (2007). *The development of national educational standards: An expertise*. Bonn: BMBF.
- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3 – 22). Cambridge, MA: Hogrefe & Huber.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG [Competence models for assessing individual learning outcomes and evaluating educational processes: Description of a newly established GRF priority program]. *Zeitschrift für Pädagogik [Journal of pedagogy]*, 52(6), 876–903.
- Krell, M. (2012). Using polytomous IRT models to evaluate theoretical levels of understanding models and modeling in biology education. *Science Education Review Letters, Theoretical Letters*, 2012, 1–5. Retrieved from edoc-server. (urn:nbn:de:kobv:11-100205516).
- Krell, M., Upmeyer zu Belzen, A., & Kruger, D. (2012). Students' understanding of the purpose of models in different biological contexts. *International Journal of Biology Education*, 2, 1 – 34. Retrieved from http://www.ijobed.com/2_2/Moritz-2012.pdf
- Krell, M., Upmeyer zu Belzen, A., & Kruger, D. (2013). Students' levels of understanding models and modelling in biology: Global or aspect-dependent? *Research in science education*, 1 – 24. doi:10.1007/s11165-013-9365-y
- Leutner, D., Fleischer, J., Spoden, C., & Wirth, J. (2007). Landesweite Lernstandserhebungen zwischen Bildungsmonitoring und Individualdiagnostik [State-wide standardised assessments of learning between educational monitoring and individual diagnostics] [special issue]. *Zeitschrift für Erziehungswissenschaften [Journal for educational science]*, 8, 149–167.
- Louca, T. L., & Zacharia, C. Z. (2012). Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, 64(4), 471–492. doi:10.1080/00131911.2011.628748
- Louca, T. L., Zacharia, Z., Michael, M., & Constantinou, P. C. (2011). Objects, entities, behaviors and interactions: A typology of student-constructed computer-based models of physical phenomena. *Journal of Educational Computing Research*, 44(2), 173–201. doi:10.2190/EC.44.
- Mahr, B. (2009). Information science and the logic of models. *Software & Systems Modeling*, 8, 365 – 383.

- MAXQDA (Version 10) [Computer Software]. Berlin: VERBI Software. Retrieved from <http://www.maxqda.de/>
- Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* [Qualitative analysis: Basics and techniques] (11th, updated and revised ed.). Weinheim: Beltz.
- Mikelskis-Seifert, S., & Fischler, H. (2003). Die Bedeutung des Denkens in Modellen bei der Entwicklung von Teilchenvorstellungen—Stand der Forschung und Entwurf einer Unterrichtskonzeption [The relevance of thinking in models with regard to conceptions of particles—state of research and draft of a teaching conception]. *Zeitschrift für die Didaktik der Naturwissenschaften* [Journal of science education], 9, 75–88.
- National Research Council. (2000). *Inquiry and the national science education standards*. Washington, DC: National Academy Press.
- Odenbaugh, J. (2005). Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology. *Biology & Philosophy*, 20(2–3), 231–255. doi:10.1007/s10539-004-0478-6
- Oh, P. S., & Oh, S. J. (2011). What teachers of science need to know about models: An overview. *International Journal of Science Education*, 33(8), 1109–1130. doi:10.1080/09500693.2010.502191
- Rost, J. (2004). *Lehrbuch Testtheorie—Testkonstruktion* [Textbook test theory—test construction]. Bern: Huber.
- Schwarz, C., Reiser, B., 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. Current challenges and future directions* (pp. 101 – 137). Rotterdam: Sense.
- Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., ..., Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632 – 654. doi:10.1002/tea.20311
- Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, 23(2), 165 – 203. doi:10.1207/s1532690xci2302_1
- Sins, P. H. M., Savelsbergh, E. R., Joolingen, W. R. van, & Hout-Wolters, B. H. A. M. van (2009). The relation between students' epistemological understanding of computer models and their cognitive processing on a modelling task. *International Journal of Science Education*, 31(9), 1205 – 1229. doi:10.1080/09500690802192181
- Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (Eds.). (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004* [Educational standards for the subject biology for intermediate-level education. By order of 16 December 2004]. München: Wolters Kluwer.
- Terzer, E. (2013). *Modellkompetenz im Kontext Biologieunterricht—Empirische Beschreibung von Modellkompetenz mithilfe von Multiple-Choice Items* [Model competence in the context of biology education—empirical description of model competence using multiple-choice items]. (Doctoral dissertation). Retrieved from edoc-server. (urn:nbn:de:kobv:11-100206767).
- Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, 24(4), 357 – 368. doi:10.1080/09500690110066485
- Trier, U., & Upmeyer zu Belzen, A. (2009). 'Die Wissenschaftler nutzen Modelle, um etwas Neues zu entdecken, und in der Schule lernt man einfach nur, dass es so ist.': Schülervorstellungen zu Modellen ['Scientists use models to discover something new and in school, you only learn that this is the case': Students' conceptions of models]. In D. Krüger, A. Upmeyer zu Belzen, S. Hof, K. Kremer, & J. Mayer (Eds.), *Erkenntnisweg Biologiedidaktik 8* [Path of knowledge biology education 8] (pp. 23–37). Kassel: Universität Kassel.

- Upmeyer zu Belzen, A., & Kruger, D. (2010). Modellkompetenz im Biologieunterricht [Model competence in biology education]. *Zeitschrift für Didaktik der Naturwissenschaften* [Journal of science education], 16, 41–57.
- Van Der Valk, T., Van Driel, J., & De Vos, W. (2007). Common characteristics of models in present-day scientific practice. *Research in Science Education*, 37(4), 469–488. doi: 10.1007/s11165-006-9036-3
- Van Driel, J. H., & Verloop, N. (1999). Teachers' knowledge of models and modelling in Science. *International Journal of Science Education*, 21(11), 1141–1153. doi:10.1080/095006999290110
- Weinert, F. E. (2001). Concepts of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45 – 65). Seattle, WA: Hogrefe & Huber.

Beitrag 4

Grünkorn, J., Lotz, A. & Terzer, E. (2014). Erfassung von Modellkompetenz im Biologieunterricht. *Der mathematische und naturwissenschaftliche Unterricht*, **67** (3), 132-138.

Erfassung von Modellkompetenz im Biologieunterricht

JULIANE GRÜNKORN – ALEXANDER LOTZ – EVA TERZER

Grundlage für kompetenzorientiertes Unterrichten sind Informationen über den aktuellen Kompetenzstand der Schüler. Für Modellkompetenz im Biologieunterricht schlägt der vorliegende Beitrag ein Raster zur Selbsteinschätzung für Lernende und einen Diagnosebogen vor, den Lehrkräfte zur Erfassung und Einordnung von Schüleraussagen anhand eines Kategoriensystems nutzen können.

1 Einleitung

Das Ziel naturwissenschaftlichen Unterrichts, Fachwissen zu vermitteln, kann erreicht werden, wenn Modelle als Medien eingesetzt werden (HENZE, VAN DRIEL & VERLOOP, 2007). Lehrkräfte können mit Modellen demonstrieren, wie etwas funktioniert, und komplexe wissenschaftliche Konzepte anschaulich erklären, so dass sie für Schüler zugänglich sind (OH & OH, 2011). Mit dem Einsatz von Modellen als Medien wird ihr Potential jedoch nicht ausgeschöpft. Vor allem die Entwicklung und Anwendung von Modellen bieten weitere, vielfältige Lerngelegenheiten (MORRISON & MORGAN, 1999). Sowohl mit der Entwicklung als auch der Anwendung von Modellen sind spezifische Lernprozesse im Bereich der Erkenntnisgewinnung verbunden, die im naturwissenschaftlichen Unterricht dazu beitragen, etwas über die Naturwissenschaften und ihre Methoden zu erfahren (HENZE et al., 2007). Der vorliegende Beitrag soll helfen, diese Kompetenzen zu entwickeln. Dazu stellt dieser Beitrag ein erprobtes Raster vor, das Lernenden lernbegleitend Rückmeldungen zu ihrem Kompetenzstand gibt und sie auf diese Weise darin unterstützt, ihren Lernprozess zielgerichtet zu strukturieren. Darüber hinaus wird gezeigt, wie durch den Einsatz eines Diagnosebogens die Modellkompetenz von Lernenden auf alltagstaugliche Art und Weise erfasst und mithilfe eines Kategoriensystems zur Modellkompetenz eingeordnet werden kann.

2 Schülerperspektiven auf Modelle und Modellbildung im Kontext Biologieunterricht

Das Kompetenzmodell der Modellkompetenz von UPMEIER ZU BELZEN und KRÜGER (2010) bildet einen Ausgangspunkt dafür, Ziele für den Umgang mit Modellen zu konkretisieren, Diagnosen einzuordnen und gezielte Fördermaßnahmen zur Entwicklung von Modellkompetenz im Biologieunterricht zu planen. Dieses Kompetenzmodell umfasst die Teilkompetenzen ‚Eigenschaften von Modellen‘, ‚Alternative Modelle‘, ‚Zweck von Modellen‘, ‚Testen von Modellen‘ und ‚Ändern von Modellen‘. Diese Inhalte graduieren UPMEIER ZU BELZEN und KRÜGER (2010) in drei Niveaus. Niveau I und II bilden eine mediale Perspektive auf Modelle mit einem naiv-realistischen bzw. relativistischen Wissenschaftsverständnis ab. Niveau III bezieht sich auf die methodische Perspektive auf Modelle und ein konstruktivistisch geprägtes Wissenschaftsverständnis. Damit schafft das Kompetenzmodell Anknüpfungspunkte für eine Erfassung von bereits verfügbaren Kompetenzen der Schüler sowie für eine gezielte Förderung.

2.1 Kategoriensystem zur Einordnung von Schüleraussagen

GRÜNKORN, UPMEIER ZU BELZEN und KRÜGER (2014) haben verschiedene Perspektiven auf Modelle und Modellbildung (Tab. 1-5) in einer Studie mit 1177 Lernenden (7.-10. Jahrgangsstufe, Gymnasien) in Berlin erhoben. Diese Studie hat Aufgaben im offenen Antwortformat eingesetzt, die sich auf verschiedene biologische Situationen beziehen. Für die Auswertung der Aufgaben wurde ein Kategoriensystem entwickelt, welches verschiedene Perspektiven von Lernenden auf Modelle und Modellbildung als Kategorien beschreibt (GRÜNKORN et al., 2014). Diese Kategorien sind innerhalb der jeweiligen Teilkompetenz nach steigender Komplexität strukturiert (Tab. 1-5). Das hier gezeigte Kategoriensystem (Tab. 1-5) ist soweit für schulische Zwecke reduziert, dass mit großer Wahrscheinlichkeit mit Aussagen in den präsentierten Kategorien gerechnet werden kann. Lehrkräfte können dieses Kategoriensystem dazu nutzen, sich grundsätzlich mit verschiedenen Schülerperspektiven über Modelle und Modellbildung auseinanderzusetzen. Darüber hinaus dient es dazu, das Niveau von Schüleraussagen einzuschätzen.

2.2 Zentrale Schülerperspektiven

Die Ergebnisse von GRÜNKORN et al. (2014) zeigen, dass Schüler im Gegensatz zu wissenschaftlich Tätigen Modelle vorwiegend als Medien und weniger als Mittel zur Erkenntnisgewinnung verstehen. Die hier beschriebenen Befunde zu Vorstellungen von Lernenden stimmen mit anderen nationalen (z. B. TERZER, PATZKE & UPMEIER ZU BELZEN, 2012, TRIER & UPMEIER ZU BELZEN, 2009) und internationalen Studien (z. B. CAREY, EVANS, HONDA, JAY & UNGER, 1989, GROSSLIGHT, UNGER, JAY & SMITH, 1991) überein und zeigen, dass ein Förderbedarf besteht.

2.2.1 Eigenschaften von Modellen

Verbreitete Schülervorstellungen zur Frage, inwiefern die gezeigten Modelle mit dem Original vergleichbar wären, sind, dass Modelle maßstabsgetreue Kopien der Originale darstellen (Tab. 1. ‚Modell als Kopie‘) oder dass Modelle den Originalen sehr ähnlich sehen (Tab. 1. ‚Modell mit großer Ähnlichkeit‘). Lernende begründen diese Vorstellung häufig mit einem großen Vertrauen gegenüber der Arbeit von Wissenschaftlern: »...weil bestimmt viele Biologen an diesem Modell gearbeitet haben. Diese Leute kennen sich bestimmt mit so was aus, also glaube ich

Kategorienname	Beispielantworten
<i>Niveau I</i>	
Modell als Kopie	Das Modell sieht genauso aus wie das Original und gleicht ihm in allen Eigenschaften. Das Modell ist eine vergrößerte/verkleinerte Kopie des Originals.
Modell mit großer Ähnlichkeit	Das Modell ähnelt dem Original in fast allen Eigenschaften. Bestimmte Eigenschaften des Originals sind vom Modellierer nicht gut umgesetzt worden oder entsprechen nicht meinen Vorstellungen.
<i>Niveau II</i>	
Modell ist in Teilen eine Kopie	Das Modell gleicht nur in bestimmten Merkmalen dem Original. Andere Merkmale sind für den Modellierer nicht bekannt und können daher nicht beurteilt werden.
Modell als fokussierte Darstellung	Das Modell ist vereinfacht, d. h. nur bestimmte Merkmale sind dargestellt und hervorgehoben. Andere Merkmale sind nicht gezeigt.
<i>Niveau III</i>	
Modell als hypothetische Darstellung	Das Modell stellt eine Idee/Hypothese über das Original dar. Eine Ähnlichkeit zwischen Original und Modell ist möglich.

Tab. 1. Schülerperspektiven der Teilkompetenz ‚Eigenschaften von Modellen‘

schon, dass er [der Neandertaler] so aussah« (Schülerin, 7. Klasse). Weniger verstehen Lernende Modelle als idealisierte Repräsentationen, die bestimmte Merkmale oder Eigenschaften des Originals hervorheben oder sehen Modelle als hypothetische Rekonstruktionen, denen Hypothesen zugrunde liegen (Tab. 1).

2.2.2 Alternative Modelle

Mit Bezug auf die Teilkompetenz ‚Alternative Modelle‘ begründen Lernende die Existenz verschiedener Modelle zu einem Original meistens damit, dass Modelle unterschiedliche Eigenschaften haben, wie beispielsweise unterschiedliche Farben oder aus unterschiedlichen Materialien bestehen (Tab. 2, ‚Unterschiedliche Modellobjekteigenschaften‘). An dieser Stel-

le beziehen sich Lernende nur auf das gegenständliche Modell und lassen die Hypothese über das Original, die diesem Modell zugrunde liegt, außer Acht. Viele Lernende begründen alternative Modelle zu einem Original aber auch mit unterschiedlichen inhaltlichen Schwerpunkten/Blickwinkeln (Struktur oder Funktion, Innen- oder Außenansicht), die der Modellierer präsentieren möchte. Eine Schülerin (10. Klasse) schreibt: »Da in einem Modell häufig nicht alle Aspekte berücksichtigt werden können, ... muss man mehrere Modelle erstellen. Im Modell A wird beispielsweise der Aufbau vermittelt...«. Hier stellt die Schülerin einen Bezug zum Original her, erkennt die Komplexität des Originals an und versteht, dass mit einem Modell nicht alle Aspekte eines Originals repräsentiert werden können (Tab. 2, ‚Unterschiedliche inhaltliche Schwerpunkte‘). Die Perspektive, dass Modelle verschiedene Hypothesen über das Original darstellen, zeigen Lernende seltener (Tab. 2).

Kategorienname	Beispielantworten
<i>Basales Niveau</i>	
Verschiedene Modelle zu unterschiedlichen Originalen	Es gibt verschiedene Modelle zu einem Original, weil jedes Modell ein anderes Original repräsentiert.
Nur ein endgültiges und richtiges Modell	Es gibt verschiedene Modelle zu einem Original, weil nur eines der alternativen Modelle korrekt ist und die anderen falsch sind. Es gibt verschiedene Modelle zu einem Original, weil diese Modelle den historischen Entwicklungsprozess zeigen sollen und nur eines dieser Modelle das endgültige und richtige Modell ist.
<i>Niveau I</i>	
Unterschiedliche Modellobjekteigenschaften	Es gibt verschiedene Modelle zu einem Original, weil die Modelle unterschiedlich gebaut sind (2-D/3-D, verschiedene Farben, verschiedene Materialien). Es gibt verschiedene Modelle zu einem Original, weil die Modelle unterschiedliche Eigenschaften haben (beweglich/unbeweglich, weich/hart).
<i>Niveau II</i>	
Unterschiedliche inhaltliche Schwerpunkte	Es gibt verschiedene Modelle zu einem Original, weil das Original komplex ist und verschiedene Modelle notwendig sind, um unterschiedliche Perspektiven auf das Original zu zeigen (u. a. innen/außen, Längs- /Querschnitt, Struktur/Funktion, verschiedene Ausschnitte bzw. Zustände des Originals).
<i>Niveau III</i>	
Unterschiedliche Annahmen	Es gibt verschiedene Modelle zu einem Original, weil aus den Untersuchungen am Original unterschiedliche Vermutungen/Hypothesen/Ideen abgeleitet werden können.

Tab. 2. Schülerperspektiven der Teilkompetenz ‚Alternative Modelle‘

Kategorienname	Beispielantworten
<i>Niveau I</i>	
Modell zum Darstellen eines Sachverhaltes	Das Modell hat den Zweck, Merkmale/Sachverhalte des Originals darzustellen.
<i>Niveau II</i>	
Modell zum Erkennen/Erklären von Zusammenhängen	Das Modell hat den Zweck, Zusammenhänge zwischen verschiedenen Aspekten im Original zu erklären und bekannte Tatsachen nachzuvollziehen.
<i>Niveau III</i>	
Modell zum Überprüfen von Ideen	Das Modell hat den Zweck, Voraussagen über das Original abzuleiten. Das Modell hat den Zweck, Hypothesen über das Original zu prüfen und Schlüsse über das Original zu ziehen. Das Modell hat den Zweck, Erkenntnisse über das Original auf andere Phänomene zu übertragen.

Tab. 3. Schülerperspektiven der Teilkompetenz ‚Zweck von Modellen‘

2.2.3 Zweck von Modellen

Den Zweck von Modellen sehen Lernende darin, Merkmale des Originals (Tab. 3. ‚Modell zum Darstellen eines Sachverhalts‘) zu veranschaulichen. So schreibt ein Schüler (10. Klasse), dass der Zweck eines gezeigten Wald-Modells darin besteht, »die verschiedenen Pflanzen, die [in] einem Wald vorkommen« zu veranschaulichen. Dieser Schüler beschreibt nicht, dass mit dem Modell auch Zusammenhänge oder Prozesse erklärt werden können. Diese Perspektive (Tab. 3. ‚Modell zum Erkennen/Erklären von Zusammenhängen‘) ist ebenfalls prominent unter Lernenden, wie eine Schülerin (8. Klasse) zeigt: »... Außerdem wird auch erklärt, dass das [Meer-]Modell eine Kreislaufkette ist. Also die Pflanzen können nicht ohne den Sauerstoff und das Wasser, die Fische können nicht ohne die Pflanzen leben.« Auch hier zeigt sich, dass Modelle eher als Medien zur Veranschaulichung verstanden werden. Wenige Lernende erkennen, dass Modelle zudem als Mittel zur Erkenntnisgewinnung geeignet sind, mit denen Ideen bzw. Hypothesen untersucht werden können, um neue Erkenntnisse über das Original zu gewinnen (Tab. 3).

2.2.4 Testen von Modellen

Bei der Frage, wie Modelle getestet werden, überprüfen Schüler Modelle häufig aus einer medialen Perspektive heraus, indem sie die Ähnlichkeit zwischen Modell und Original vergleichen (Tab. 4. ‚Vergleich zwischen Modell und Original‘) und deren Übereinstimmung bewerten (Tab. 4. ‚Vergleich und Passung zwischen Modell und Original‘). Einige Lernende überprüfen zudem am gegenständlichen Modell, ob das Material geeignet ist (Tab. 4. ‚Überprüfung des Materials‘) oder ob technische Anforderungen für den Einsatz des Modells erfüllt sind (Tab. 4. ‚Überprüfung der Grundvoraussetzungen‘). Selten nutzen Lernende Modelle, um Ideen oder Hypothesen zu überprüfen (Tab. 4). Bei Schülerantworten lässt sich des Öfteren eine Schritt-für-Schritt-Prozedur erkennen, die verschiedene Perspektiven und Niveaus widerspiegeln, wie eine Schülerin (8. Klasse) zeigt: »Als erstes überprüft man, ob das Modell überhaupt fliegen kann. Dann schaut man, welche Oberflächenstruktur vorhanden ist und vergleicht sie mit dem Modell. Danach schaut man, ob alle Bestandteile vorhanden sind. Wenn das O.K. ist, stellt

Kategorienname	Beispielantworten
<i>Basales Niveau</i>	
Keine Testung des Modells	Eine Testung des Modells bzw. mit dem Modell ist unnötig.
<i>Niveau I</i>	
Überprüfung des Materials	Das Modell wird getestet, indem das Material auf Widerstandsfähigkeit (u. a. Beweglichkeit, Stabilität, Elastizität) überprüft wird.
Überprüfung der Grundvoraussetzungen	Das Modell wird getestet, indem geprüft wird, ob technische Anforderungen (z. B. das Fliegen bei einem Vogel-Modell) für den Einsatz des Modells erfüllt sind.
<i>Niveau II</i>	
Vergleich zwischen Original und Modell	Das Modell wird getestet, indem es mit dem Original verglichen (Struktur und/oder Funktion) wird.
Vergleich und Passung zwischen Original und Modell	Das Modell wird getestet, indem geprüft wird, ob das Modell in den notwendigen Merkmalen (Struktur und/oder Funktion) mit dem Original übereinstimmt.
<i>Niveau III</i>	
Überprüfung von Hypothesen	Das Modell wird angewendet, um in einer Untersuchung mit dem Modell, eine Hypothese über das Original zu überprüfen und Erkenntnisse über das Original zu gewinnen.

Tab. 4. Schülerperspektiven der Teilkompetenz ‚Testen von Modellen‘

man Vermutungen über die Orientierung des Käfers an...«. Diese Schülerin überprüft zunächst, ob die Voraussetzungen für den Einsatz des Käfer-Modells erfüllt sind (Niveau I, Kategorie ‚Überprüfung der Grundvoraussetzungen‘). Ein bewertender Vergleich des Modells mit dem Original erfolgt im zweiten und dritten Satz (Niveau II, Kategorie ‚Vergleich und Passung zwischen Original und Modell‘). Die Aussage der Schülerin endet mit dem Hinweis, dass mit dem Modell Vermutungen über das Original geprüft werden (Niveau III, Kategorie ‚Überprüfung von Hypothesen‘).

2.2.5 Ändern von Modellen

Eine prominente Schülerperspektive für die Teilkompetenz ‚Ändern von Modellen‘ ist, dass Modelle geändert werden, wenn sie in Struktur und/oder Funktion nicht mit dem Original übereinstimmen (Tab. 5, ‚Ändern bei mangelnder Passung mit dem Original‘). Bei Lernenden zählt an dieser Stelle eine genaue Übereinstimmung zwischen Original und Modell ungeachtet des Zwecks des Modells, für den es gebaut wurde. Der Zweck ist jedoch zentral, um Modelle angemessen zu beurteilen. Eine weitere zentrale Perspektive ist, dass Schüler Modelle ändern, um sie technisch besser auszustatten oder um sie funktionsfähiger zu machen. Das Original als auch der Zweck des Modells bleiben hier unbeachtet (Tab. 5, ‚Ändern zur Verbesserung des Modellobjekts‘). Wie bei der Teilkompetenz ‚Testen von Modellen‘ ist auch hier eine Schritt-für-Schritt-Prozedur bei Lernenden zu erkennen. Ein Schüler (7. Klasse) formuliert, dass das gezeigte Libellen-Modell dann verändert werden muss, »wenn es erst gar nicht fliegt oder wenn es nicht wie eine Libelle fliegen kann«. Für den Lernenden ist somit eine Änderung am Modell notwendig, wenn (a) die Voraussetzungen für den Einsatz des Modell nicht erfüllt sind (Niveau I, Kategorie ‚Ändern bei Nichterfüllung der Grundvoraussetzungen‘) und wenn (b) das Modell nicht mit dem Original übereinstimmt (Niveau II, Kategorie ‚Ändern bei mangelnder Passung mit dem Original‘).

Insgesamt zeigen die Ergebnisse, dass nur wenige Lernende Modelle als Mittel zur Erkenntnisgewinnung in Betracht ziehen, so dass diese geändert werden, weil sich Hypothesen, die mit ihrer Hilfe aufgestellt und untersucht werden konnten, als nicht haltbar erwiesen haben (Tab. 5).

2.2.6 Basale Perspektiven

Neben diesen zentralen Perspektiven auf Modelle und Modellbildung gibt es einige Schüler, die basale Perspektiven (Tab. 2, 4 und 5) zeigen, die für die Unterrichtsplanung ebenfalls relevant sind. So bezweifeln einige Lernende die Existenz alternativer Modelle zu einem Original. Sie begründen beispielsweise verschiedene Modelle zu einem Original damit, dass diese Modelle unterschiedliche Originale repräsentieren, obwohl deutlich hervorgehoben wird, dass es Modelle zu einem Original sind (Tab. 2, ‚Verschiedene Modelle zu unterschiedlichen Originalen‘). Andere beurteilen Modelle nach ihrer Richtigkeit. Wiederum andere begründen eine Vielzahl von Modellen zu einem Original mit einer historischen Abfolge, von denen aber nur eines der Modelle das endgültige und richtige Modell ist (Tab. 2, ‚Nur ein endgültiges und richtiges Modell‘). So begründet ein Schüler (9. Klasse) die Existenz alternativer Modelle zur Biomembran wie folgt: »Ich denke, dass zwei Modelle sehr alte Modelle sind. Zu dieser Zeit hatte man noch nicht alle Informationen, wie man sie heute hat. Die stimmen nicht mehr. Das eine Modell ist aber das endgültige Modell, was auch stimmt. Ich habe es in meinem Buch gesehen und auch so gelernt.« Diese Aussage zeigt, dass ein reflektierter Umgang mit historischen Modellen im Unterricht wichtig ist, sodass nicht der Eindruck entsteht, dass Modelle endgültig sind, sobald sie einen längeren Prozess durchlaufen haben und in Büchern dokumentiert sind. Auch für die Teilkompetenzen ‚Testen‘ und ‚Ändern von Modellen‘ lassen sich basale Perspektiven beschreiben, in denen Lernende ein Testen (Tab. 4, ‚Keine Testung des Modells‘) und Ändern von Modellen (Tab. 5, ‚Kein Anlass für eine Änderung‘) ablehnen.

Kategorienname	Beispielantworten
<i>Basales Niveau</i>	
Kein Anlass für eine Änderung	Modelle sind endgültig und werden nicht verändert.
Ändern zur Darstellung unterschiedlicher Originale	Modelle werden verändert, weil es unterschiedliche Originale gibt und jedes Original durch ein Modell repräsentiert werden muss.
<i>Niveau I</i>	
Ändern zur Verbesserung des Modellobjekts	Modelle werden verändert, damit das Modell funktionsfähiger (z. B. bessere Technik) und ästhetischer wird.
Ändern bei Nichterfüllung der Grundvoraussetzungen	Modelle werden verändert, um Materialfehler im Modell zu beheben. Modelle werden verändert, wenn sie nicht die technischen Anforderungen (z. B. das Fliegen bei einem Vogel-Modell) für den Einsatz erfüllen.
<i>Niveau II</i>	
Ändern bei mangelnder Passung mit dem Original	Modelle werden verändert, wenn sie nicht mit dem Original übereinstimmen (Struktur und/oder Funktion).
Ändern bei neuen Erkenntnissen über das Original	Modelle werden verändert, wenn neue Informationen/Erkenntnisse über das Original im Modell berücksichtigt werden sollen.
<i>Niveau III</i>	
Ändern bei Erkenntnissen aus Modellexperimenten	Modelle werden verändert, wenn auf der Grundlage von Ergebnisse aus einem Experiment mit Modellen, die Hypothese über das Original falsifiziert wurde.

Tab. 5. Schülerperspektiven der Teilkompetenz ‚Ändern von Modellen‘

In den kommenden Wochen werden Sie im Biologieunterricht mit Modellen arbeiten. Beenden Sie folgende Satzanfänge möglichst ausführlich. Sollten Sie einen Satzanfang nicht beenden können, begründen Sie bitte möglichst ausführlich, warum nicht. Ihre Antworten werden nicht bewertet.

Die Beziehung zwischen Modell und Original besteht darin, dass ...

Mehrere Modelle zu einem Original sind sinnvoll, weil ...

Modelle dienen dazu, dass ...

Modelle überprüft man, indem ...

Modelle werden verändert, weil ...

Abb. 1. Diagnosebogen

3 Instrumente zur Erfassung von Modellkompetenz in der Schule

Zum einen wurde ein Diagnosebogen konzipiert (Abb. 1), der Vorstellungen zu Modellen erfasst. Hierzu beenden die Schüler vorgegebene Satzanfänge (vgl. GROPPENGIESSER, 1996), die sich auf die einzelnen Teilkompetenzen von Modellkompetenz beziehen. Die Satzanfänge sind in Anlehnung an Interviewfragen von GROSSLIGHT et al., (1991) formuliert, die ebenfalls Schüler zu Modellen befragt haben. Die Antworten auf den Diagnosebogen können mit Hilfe des vorgestellten Kategoriensystems (Tab. 1-5) ausgewertet werden, sodass eine Einschätzung der Modellkompetenz möglich ist (vgl. KUNTER & KLUSMANN, 2010).

Zum anderen wurde ein Kompetenzraster entwickelt (Tab. 6), mit dem Schüler ihre Modellkompetenz mit Blick auf ein konkretes Unterrichtsthema selbst einschätzen können. Gleichzeitig erfahren sie bei der Arbeit mit dem Kompetenzraster, welches Entwicklungspotential sie noch haben. Die Formulierung der »Ich kann«-Beschreibungen beruht auf dem Kategoriensystem von GRÜNKORN et al. (2014), der Operationalisierung von Modellkompetenz in Multiple-Choice Items (TERZER, HARTIG & UPMEIER ZU BELZEN, 2013) sowie Vorstellungen von Lernenden, die diese während der Bearbeitung von Multiple-Choice Items mit der Methode des lauten Denkens geäußert haben (TERZER et al., 2012). Tabelle 6 zeigt allgemein gehaltene Formulierungen, die auf den jeweiligen Unterrichtsinhalt angepasst werden. Die Aussage »Ich kann Eigenschaften beschreiben, in denen Modell und Original sich gleichen« kann z. B. in Bezug auf eine Unterrichtseinheit zur Biomembran spezifisch als »Ich kann Eigenschaften beschreiben, in denen Fluid-Mosaik-Modell und Biomembran sich gleichen« formuliert werden. Das Kompetenzraster bietet die Möglichkeit, den Lernenden Rückmeldungen über den Lernstand zu geben. Auf diese Weise können subjektive Sichtweisen bzw. Selbstkonzepte der Lernenden dokumentiert werden (vgl. KUNTER & KLUSMANN, 2010).

4 Einsatz der Instrumente

Ein Thema im Biologieunterricht, in dem Modelle eine zentrale Rolle spielen, ist die Biomembran. Im Forschungsprozess dienen Modelle zur Biomembran dazu, unterschiedliche Vorstellungen zu präsentieren und tragen gleichzeitig dazu bei, neues Wissen über die Biomembran zu generieren. Hierzu gibt es bereits in Schulbüchern eine Fülle von bewährten Unterrichtsmaterialien. Aus diesem Grund wurden die hier vorgestellten Instrumente in einer Unterrichtseinheit zum Thema Biomem-

bran in der 12. Jahrgangsstufe mehrfach eingesetzt und auf ihre Praktikabilität getestet.

Vor Beginn der Unterrichtseinheit füllen die Lernenden den Diagnosebogen aus, so dass die Gestaltung des Unterrichts auf ihre verfügbare Modellkompetenz abgestimmt werden kann. Da die Auswertung des Diagnosebogens anhand des Kategoriensystems (Tab. 1-5) von der Lehrkraft vorgenommen und die Unterrichtseinheit nach den Einschätzungen darauf abgestimmt wird, ist zu empfehlen, hierfür genug Zeit einzuplanen. Vor dem Einsatz des Diagnosebogens sollte die Lehrkraft den Lernenden dessen Intention transparent machen. Diese besteht darin, die erfassten Schülerperspektiven zu nutzen, um den Unterricht passgenauer zu planen und nicht um die Schüler zu bewerten.

Die Vorstellungen, die Lernende in ihren Antworten auf den Diagnosebogen nennen, bewegen sich häufig auf Niveau I und decken sich in ihren Schwerpunkten mit denen, die im Abschnitt 2.2 beschrieben werden. Schwierigkeiten treten zum Teil in der Beendigung von Satzanfängen auf. Dies ist insbesondere bei der Teilkompetenz ‚Eigenschaften von Modellen‘ und vereinzelt bei der Teilkompetenz ‚Ändern von Modellen‘ zu beobachten. Lernende, die diese Probleme haben, zeigen in den Formulierungen der anderen Teilkompetenzen häufig eine geringe Ausprägung von Modellkompetenz. Schwierigkeiten bzw. Unsicherheiten seitens der Schüler, Erklärungen für Änderungsprozesse bei Modellen zu formulieren, beschreiben auch Studien von GROSSLIGHT et al. (1991) und GRÜNKORN et al. (2014), in denen Lernende z. T. sehr vage Erklärungen geben.

Die Ergebnisse des Diagnosebogens kann die Lehrkraft im Unterricht als prominente Vorstellungen zusammenfassen und diese der wissenschaftlichen Sicht auf Modelle als Mittel der Erkenntnisgewinnung gegenüberstellen. An dieser Stelle ist von wertenden Formulierungen abzusehen. Zudem ist es sinnvoll, die Diagnosebögen namentlich oder in anonymisierter Form zu kennzeichnen, so dass im Laufe des Unterrichts der individuelle Kompetenzfortschritt dokumentiert werden kann. Ein anonymisiertes Vorgehen ist dann sinnvoll, wenn die Lernenden den unbewerteten Einsatz von Diagnoseinstrumenten noch nicht gewohnt sind. Damit ist es möglich, Hemmungen oder Ängsten, beim Beantworten »etwas falsch zu machen« zu begegnen.

Darüber hinaus schätzen sich die Schüler vor Beginn der Unterrichtseinheit in einem Kompetenzraster ein, welches die Erwartungen an die Lernenden transparent macht und der Lehrkraft Einblicke in die Selbstkonzepte der Schüler zu Modellkompetenz gibt. Indem die Lernenden die Ergebnisse von Diagnosebogen und Kompetenzraster vergleichen, lernen sie neben der fachlichen Auseinandersetzung mit Modellen, ihre

Teilkompetenzen	Lernfortschritt		
	Ich kann Eigenschaften beschreiben, in denen Modell und Original sich unterscheiden.	Ich kann Eigenschaften beschreiben, in denen Modell vereinfacht dargestellt sind.	Ich kann die jeweilige Hypothese formulieren, die einem bestimmten Modell zu Grunde liegt.
Eigenschaften von Modellen	Ich kann beschreiben, inwiefern sich verschiedene Modelle in ihren gegensätzlichen Eigenschaften (z. B. Farben, Materialien) unterscheiden.	Ich kann beschreiben, inwiefern sich verschiedene Modelle in ihren inhaltlichen Schwerpunkten (z. B. Funktion/Struktur, innen/außen, Längs-/ Querschnitt) unterscheiden.	Ich kann die jeweiligen Hypothesen formulieren, die unterschiedlichen Modellen zu Grunde liegen.
Alternative Modelle	Ich kann mit dem Modell die Merkmale des Originals veranschaulichen.	Ich kann mit dem Modell Zusammenhänge im Original erklären.	Ich kann mit dem Modell Hypothesen über das Original ableiten.
Zweck von Modellen	Ich kann beschreiben, inwiefern sich verschiedene Modelle in ihren gegensätzlichen Eigenschaften (z. B. Farben, Materialien) unterscheiden.	Ich kann mit dem Modell die Merkmale des Originals veranschaulichen.	Ich kann mit dem Modell Hypothesen über das Original ableiten.
Testen von Modellen	Ich kann das Modell testen, indem ich das Material auf Widerstandsfähigkeit überprüfe.	Ich kann das Modell testen, indem ich überprüfe, ob technische Anforderungen für den Einsatz (z. B. das Fliegen bei einem Vogelmodell) erfüllt sind.	Ich kann ein Modell anwenden, um mit dem Modell eine Hypothese über das Original zu überprüfen und Erkenntnisse über das Original zu gewinnen.
Ändern von Modellen	Ich kann ein Modell so verändern, dass Fehler im Modell (z. B. Materialmängel) verbessert sind.	Ich kann ein Modell so verändern, dass neue Informationen über das Original in das Modell einbezogen sind.	Ich kann ein Modell auf der Grundlage von Ergebnissen aus einem Experiment mit dem Modell so verändern, dass es zu einer daraufhin veränderten Hypothese über das Original passt.

Tab. 6. Kompetenzraster zur Modellkompetenz

Kompetenzen einzuschätzen und Mitverantwortung für deren Weiterentwicklung zu übernehmen. Auch hier sollte die Lehrkraft deutlich machen, dass die formulierten Erwartungen keine Bewertung einzelner darstellen, sondern eine Hilfestellung sein sollen, sich beim Lernen zu orientieren und ein Bewusstsein für den eigenen Kompetenzerwerb zu entwickeln.

Im konkreten unterrichtlichen Einsatz hat sich gezeigt, dass sich viele Schüler auf Niveau I einschätzen, was häufig mit der Fremdeinschätzung durch die Lehrkraft übereinstimmt. Unterschiede zeigen sich jedoch bei leistungsstarken Lernenden, die sich selbst auf einem geringeren Kompetenzniveau einschätzen. Derartige Unterschiede zwischen Ergebnissen von subjektiver und objektiver Kompetenzerfassung sind in anderen Studien bereits beschrieben (KUNTER & KLUSMANN, 2010). Diese Diskrepanz gibt wichtige Hinweise über das Selbstkonzept des Lernenden und kann Anlass für ein individuelles Gespräch sein. In diesen Gesprächen sind die Ursachen für diese Diskrepanz herauszufinden. Darüber hinaus sind im Sinne eines lebenslangen Lernens Hilfestellungen zu geben, das kategoriengestützte Selbsteinschätzen zu erlernen. Der kontinuierliche Einsatz derartiger Instrumente und eine wertschätzende Rückmeldung der Lehrkräfte unterstützt die Lernenden in der Entwicklung dieser Kompetenz.

Da im kompetenzorientierten Unterricht die Kompetenzen der Lernenden im Blick bleiben und systematisch entwickelt werden, sind Phasen der Erfassung und Phasen der Förderung von Kompetenzen zyklisch miteinander verbunden (SCHRÖDER & WIRTH, 2012). In der Unterrichtseinheit zur Biomembran entwickeln die Lernenden aus experimentell gewonnenen bzw. in Materialien dargestellten Daten Vorstellungen über den Bau der Biomembran, die in Modelle umgesetzt werden. Anschließend wird die Gültigkeit der Modelle an neuen Daten geprüft. Wenn sich daraus die Falsifizierung der zugrunde liegenden Hypothese ergibt, müssen die Modelle verändert werden. Auf diese Weise wird einerseits der Anspruch von Biologieunterricht eingelöst, den typischen Gang der naturwissenschaftlichen Erkenntnisgewinnung problemorientiert und hypothetisch-deduktiv zu rekonstruieren. Gleichzeitig kann durch den Einbezug der Erfassung und Reflexion der verfügbaren Modellkompetenz der Unterricht kompetenzorientiert ausgerichtet werden.

Zur Überprüfung der Lernwirksamkeit des durchgeführten Unterrichts kommt am Ende dieses Unterrichtsabschnitts der Diagnosebogen erneut zum Einsatz. Der Vergleich der individuellen Antworten gibt der Lehrkraft Einblicke in den Kompetenzzuwachs der Lernenden. Das Kompetenzraster dient auch an dieser Stelle der Orientierung in Bezug auf den Kompetenzstand. Im Sinne einer prozessorientierten Erfassung reflektieren Lernende und Lehrende den bisherigen Arbeitsverlauf. Wenn die Diagnosebögen und Kompetenzraster namentlich gekennzeichnet sind, kann die Lehrkraft bereits an dieser Stelle erste Einblicke in den Lernfortschritt einzelner Schüler bekommen. Die z. T. bei der Ersterfassung noch beobachtbaren Probleme beim Ausfüllen des Bogens tauchen an dieser Stelle erfahrungsgemäß nicht mehr auf.

Die Analyse des bereits erreichten Kompetenzstands ist Grundlage für eine gezielte Passung der inzwischen verfügbaren Kompetenzen und der Gestaltung des nächsten Unterrichtsabschnitts. In der Regel hat der Kompetenzerwerb bis zu dieser Stelle zu einer größeren Heterogenität bei den Lernenden geführt. So haben sich möglicherweise die Lernenden in einigen Teilkompetenzen weiterentwickelt, in anderen jedoch noch nicht. Deshalb werden nun von der Lehrkraft selbst entwickelte binnendifferenzierte Materialien, Aufgaben und Versuche genutzt, um für die weitere Kompetenzentwicklung passgenauer an die individuell verfügbare Modellkompetenz der Schüler anzuknüpfen.

Wichtig für den Lernprozess ist eine zusammenfassende Schlussreflexion am Ende der Unterrichtseinheit. Diese zeigt, ob und inwieweit die Schüler den Modellbildungsprozess verstanden haben, ihn nachvollziehen und anwenden können. Hierzu erstellen die Lernenden in Gruppen ein Fließschema zum Einsatz von Modellen zum Erkenntnisgewinn (vgl. FLEIGE, SEEGER, UPMEIER ZU BELZEN & KRÜGER, 2012), das von dem konkreten Unterrichtsbeispiel abstrahiert und ein übergreifendes Modellverständnis visualisiert. Die Präsentation dieser Lernprodukte bietet Anknüpfungspunkte für ein Lerngespräch über den Lernprozess und das Lernergebnis (TSCHEKAN, 2011). Es dient dazu, aus einer prozessorientierten Perspektive auf den langfristigen Kompetenzaufbau offene Fragen oder Schwierigkeiten zu verdeutlichen, die die Schüler sowie die Lehrkraft selbst in weiteren Unterrichtseinheiten bearbeiten sollten. Auf diese Weise findet ein kumulativer, vernetzter Kompetenzaufbau über verschiedene Unterrichtseinheiten hinweg statt (TSCHEKAN, 2011, SCHRÖDER & WIRTH, 2012).

Zusammenfassend hat sich in der Unterrichtspraxis gezeigt, dass die beiden hier vorgestellten Instrumente (Diagnosebogen und Kompetenzraster) probate Mittel für Lehrkräfte zur Erfassung der Kompetenzstände der Schüler darstellen. Sie unterstützen die Lehrkräfte darin, den Herausforderungen kompetenzorientierten Unterrichtens zu begegnen.

Literatur

- CAREY, S., EVANS, R., HONDA, M., JAY, E. & UNGER, C. (1989). »An experiment is when you try it and see if it works«: a study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, 11 (Special Issue), 514–529.
- FLEIGE, J., SEEGER, A., UPMEIER ZU BELZEN, A. & KRÜGER, D. (Hg.) (2012). *Modellkompetenz im Biologieunterricht Klasse 7-10*. Donauwörth: Auer.
- GROPENGIESSER, H. (1996). Die Bilder im Kopf. Von den Vorstellungen der Lernenden ausgehen. In Prüfen und Beurteilen. *Friedrich Jahresheft* 14, 11-13.
- GROSSLIGHT, L., UNGER, C., JAY, E. & SMITH, C. L. (1991). Understanding Models and their Use in Science: Conceptions of Middle and High School Students and Experts. *Journal of Research in Science Teaching*, 28 (9), 799–822.
- GRÜNKORN, J., UPMEIER ZU BELZEN, A. & KRÜGER, D. (2014). Assessing Students' Understandings of Biological Models and their Use in Science to Evaluate a Theoretical Framework. *International Journal of Science Education*. 1-34.
- HENZE, I., VAN DRIEL, J. H. & VERLOOP, N. (2007). Science teachers' knowledge about teaching models and modelling in the context of a new syllabus on public understanding of science. *Research in Science Education*, 37, 99–122.
- KUNTER, M. & KLUSMANN, U. (2010). Kompetenzmessung bei Lehrkräften – Methodische Herausforderungen. *Unterrichtswissenschaft*, 38(1), 68-86.
- MORRISON, M. & MORGAN, M. S. (1999). Introduction. In: M. S. MORGAN & M. MORRISON (Hg.): *Models as mediators. Perspectives on natural and social science*. Cambridge, UK: Cambridge University Press, 1–9.
- OH, P. S. & OH, S. J. (2011). What Teachers of Science Need to Know about Models: An overview. *International Journal of Science Education*, 33(8), 1109–1130.
- SCHRÖDER, C. & WIRTH, I. (2012). *99 Tipps. Kompetenzorientiert unterrichten*. Cornelsen: Berlin.
- TERZER, E., HARTIG, J. & UPMEIER ZU BELZEN, A. (2013). Systematische Konstruktion eines Tests zu Modellkompetenz im Biologieunterricht unter Berücksichtigung von Gütekriterien. *Zeitschrift für die Didaktik der Naturwissenschaften*, 19, 51-76.
- TERZER, E., PATZKE, C. & UPMEIER ZU BELZEN, A. (2012). Validierung von Multiple-Choice Items zur Modellkompetenz durch lautes Denken. In: U. HARMS & F. X. BOGNER (Hg.): *Lehr- und Lernforschung in der Biologiedidaktik*. Innsbruck: Studienverlag, 45–62.
- TRIER, U. & UPMEIER ZU BELZEN, A. (2009). »Wissenschaftler nutzen Modelle, um etwas Neues zu entdecken, und in der Schule lernt man einfach nur, dass es so ist.«: Schülervorstellungen zu Modellen. In: D. KRÜGER, A. UPMEIER ZU BELZEN, S. HOF, K. KREMER & J. MAYER (Hg.): *Erkenntnisweg Biologiedidaktik 8*. Kassel: Universität Kassel, 23–37.
- TSCHEKAN, K. (2011). *Kompetenzorientiert unterrichten. Eine Didaktik*. Cornelsen Scriptor: Berlin.
- UPMEIER ZU BELZEN, A. & KRÜGER, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für die Didaktik der Naturwissenschaften*, 16, 41–57.
- JULIANE GRÜNKORN, gruenkorn@dipf.de, promoviert an der Freien Universität Berlin zum Thema »Modellkompetenz im Biologieunterricht«. Derzeit arbeitet sie am Deutschen Institut für Internationale Pädagogische Forschung (DIPF) als wissenschaftliche Mitarbeiterin zur Projektkoordination im DFG-Schwerpunktprogramm »Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen«.
- ALEXANDER LOTZ, alexander.lotz.030@googlemail.com, hat an der Humboldt-Universität zu Berlin Biologie und Chemie für das Lehramt an Gymnasien studiert und sein Referendariat an einem Gymnasium in Berlin absolviert. Derzeit ist er als Studienrat am Goethe-Gymnasium in Frankfurt am Main tätig.
- Dr. EVA TERZER, eva.terzer@hu-berlin.de, hat an der Humboldt-Universität zu Berlin zur empirischen Überprüfung des Kompetenzmodells zu Modellkompetenz mit Multiple-Choice Items promoviert. Aktuell ist sie Postdoc im Forschungskolleg der Professional School of Education (PSE) an der Humboldt-Universität zu Berlin und arbeitet im Bereich der Lehrerbildung. ■

Beitrag 5

Grünkorn, J. & Hanauer, N. (2013b). Schwimmen wie Fische im Wasser – Untersuchung der Schleimschicht von Fischen. *Grundschule*, **45** (6), 28-29.

Grünkorn, J. & Hanauer, N. (2013a). Fischhaut [Materialien]. *Grundschule*, **45** (6), Beilage, XIII-XVI).

Online verfügbar unter: <http://www.die-grundschule.de>

Schwimmen wie Fische im Wasser

Untersuchung der Schleimschicht von Fischen

VON JULIANE GRÜNKORN & NICOLA HANAUER

In Zeiten von „Findet Nemo“ und Unterwasserzoos staunen Kinder über die Farben und Formen der Fische und deren Leichtigkeit, sich im Wasser scheinbar ohne großen Kraftaufwand fortzubewegen. Diese Faszination wird genutzt, um mit Modellen Erkenntnisse über das Thema „Angepasstheit von Lebewesen an ihren Lebensraum“ zu gewinnen und um über Modelle zu reflektieren.



FOTO: ICH UND DU/PIXELIOIDE

dene Fischformen existieren (z. B. Kugelfische sind rund und breit; Schollen sind flach und von oben gesehen rund; Seeforellen sind lang und schmal). Der Beitrag fokussiert auf die Eigenschaft Schleimschicht, da alle Knochenfische diese besitzen (vgl. Westheide/Rieger 2010, S. 260).

Ausgehend von dieser Eigenschaft, ergibt sich die Fragestellung, die auf dem Arbeitsblatt „Notizen zum Versuch“ (siehe S. XVI) notiert wird: Welchen Effekt hat die Schleimschicht auf die Fortbewegungseigenschaften von Fischen?

Um ein strukturiertes Vorgehen zu ermöglichen, wird im weiteren Verlauf dieses Arbeitsblatt als Arbeitsgrundlage dienen. Mit Bezug auf die Fragestellung wird eine Hypothese formuliert und auf dem Arbeitsblatt festgehalten: Die Fischhaut ist von einer rutschigen Schleimschicht überzogen, die ein Gleiten durch das Wasser verbessert.

Schwimmen wie Fische im Wasser – bei Anemonenfischen trägt die Schleimschicht dazu bei, sich schnell im Wasser fortzubewegen.

Schwimmen wie Fische im Wasser“ – mit dieser Geschichte (siehe S. XIII) können die Schülerinnen und Schüler an das Thema herangeführt werden. Sie arbeiten heraus, dass sich Fische (hier Knochenfische) im Vergleich zum Menschen mühelos und schnell im Wasser fortbewegen. Fische haben eine längliche Körperform, besitzen Schuppen und sind schwer zu greifen, da sie eine Schleimschicht

haben, die ihre Oberfläche rutschig macht. In einem Unterrichtsgespräch wird deutlich, dass die Körperform, die Schuppen und die Schleimschicht mögliche Gründe für ein müheloses und schnelles Schwimmen von Fischen sind.

An dieser Stelle muss besprochen werden, dass nicht alle Knochenfische Schuppen haben (z. B. Welse haben keine Schuppen; Barsche besitzen Schuppen) und dass verschie-

MIT ODER OHNE CREME?

Wenn die Schülerinnen und Schüler den Zusammenhang zwischen der Schleimschicht und dem verbesserten Gleiten nicht eigenständig erwähnen bzw. nicht nachvollziehen können, können die Knetmodelle, die für den Modellversuch (siehe „Bau von Fischmodellen“ auf S. XIV) gebaut werden, auch dazu genutzt werden, um diese Hypothese abzuleiten. Dafür streichen die Lernenden mit der Hand über zwei Knetmodelle (mit und ohne Creme). Sie nehmen wahr,

dass sich das Modell ohne Creme stumpf anfühlt und der Finger beim Überstreichen gestoppt wird. Demgegenüber fühlt sich das Modell mit Creme geschmeidig an, und der Finger gleitet beim Überstreichen leicht darüber hinweg. An dieser Stelle kann der Zweck von Modellen (Welchen Zweck haben die beiden Knetmodelle?) – die Voraussagekraft – thematisiert werden.

Reflexion der gebauten Knetmodelle

Die gemeinsam abgeleitete Annahme über den Effekt der Schleimschicht auf die Fortbewegungseigenschaften von Fischen dient als Ausgangspunkt für das weitere Vorgehen. Es wird deutlich, dass Daten notwendig sind, um die zuvor aufgestellte Hypothese zu überprüfen (Testen von Modellen): Die Schleimschicht verbessert das Gleiten im Wasser, wodurch sich die Fische schneller im Wasser fortbewegen können als ohne Schleimschicht.

Die Daten werden anhand eines Versuchs mit Knetmodellen zur Schleimschicht von Fischen gewonnen. Hier können der Zweck von Modellen (Welchen Zweck haben die beiden Knetmodelle?) und das Testen von Modellen (Wie können wir überprüfen, ob die gebauten Knetmodelle tauglich sind und ihren Zweck erfüllen?) besprochen werden. Es wird festgehalten, dass die Knetmodelle in diesem Versuch den Zweck haben, die genannte Hypothese zu überprüfen. Bezogen auf das Testen von Modellen (vgl. Stolperstein) wird festgehalten, dass zunächst die Knetmodelle selbst überprüft werden können, indem geschaut wird, ob diese stabil sind bzw. im Versuch nicht kaputt gehen. Weiterhin wird das Modell getestet, indem die Knetmodelle mit dem Original „Fisch bzw. Fischhaut“ verglichen werden. Vor dem Hintergrund des Zwecks der Modelle in diesem Versuch ist es besonders wichtig, dass nicht nur diese medialen Perspektiven, sondern auch eine methodische Perspektive eingenommen wird. Daher muss verdeutlicht werden, dass erst durch die Anwendung des Modells und der erhobenen Daten aus dem Versuch die Hypothese über die Schleimschicht überprüft werden kann (Modell als Methode).

WELCHES KNETMODELL GLEITET SCHNELLER?

Die Testphase (siehe „Versuchsdurchführung“ auf S. XV) soll in Gruppenarbeit (etwa vier Lernende) erfolgen. Um alle Lernenden einzubinden, können Gruppenämter vergeben werden (siehe den Beitrag von Barbara Wieder auf S. 23–25). Die Lernenden notieren sich zunächst ihre Vorüberlegungen zum Modellversuch („Notizen zum Versuch“ auf S. XVI, Aufgabe 3). Diese Überlegungen umfassen Aspekte, die die Kinder beim Versuch beachten, müssen wie beispielsweise bereits vorhandenes Wissen über Fische (z. B. längliche Form) oder Wissen zum Bau von Modellen. Vor der Durchführung muss gezielt darauf hingewiesen werden, dass die Schülerinnen und Schüler bei der Durchführung ihres Versuches sehr konzentriert und genau beobachten müssen, was passiert. Der Versuchsablauf ist sehr kurz, und es könnten ihnen sonst wichtige Beobachtungen entgehen. Aus diesem Grund wird empfohlen, einen Probeauflauf durchzuführen.

Als Ergebnis stellen die Lernenden fest, dass das Modell mit Creme schneller durch das Wasser gleitet als das Modell ohne Creme (Aufgabe 4). Die Lernenden kommen somit zu dem Schluss, dass die Schleimschicht das Gleiten im Wasser verbessert. Dies bedeutet, dass sich Fische dadurch im Wasser schnell fortbewegen können (Aufgabe 5).

WAS HABEN DIE KINDER GELERNT?

Mit den genannten Reflexionsimpulsen werden die Lernenden dazu angeregt, über den Zweck und das Testen von Modellen zu reflektieren. Es soll gesichert werden, dass ...

- dazu Modelle dienen, die Bedeutung der Schleimschicht auf die Fortbewegung der Fische vorauszusagen bzw. eine Hypothese abzuleiten und zu untersuchen (Zweck von Modellen).
- erst durch die Anwendung der Modelle und den Gewinn von Daten aus dem Modellversuch die Hypothese über die Bedeutung der Schleimschicht für die Fortbewegung überprüft werden kann (Testen von Modellen).

SCHNELL GELESEN

Welchen Effekt hat die Schleimschicht auf die Fortbewegungseigenschaften von Fischen? Die Modelle zur Schleimschicht von Fischen werden gebaut und genutzt, um eine vorher aufgestellte Annahme zu überprüfen und Schlussfolgerungen über die Fortbewegungseigenschaften von Fischen zu ziehen. Damit Lernende die Modelle als Methode nutzen, sollen sie Modelle selbst bauen, sie einsetzen und etwas über das Original erfahren.

ANGESPROCHENE KOMPETENZEN

- Zweck von Modellen: Gib an, welchen Zweck Modelle haben.
- Testen von Modellen: Erkläre, wie wir überprüfen können, ob die Modelle tauglich sind.

DIE AUTORINNEN

Juliane Grünkorn

promoviert in der Didaktik der Biologie an der Freien Universität Berlin und arbeitet derzeit am Deutschen Institut für Internationale Pädagogische Forschung in Frankfurt/Main.

Nicola Hanauer

ist ehemalige studentische Mitarbeiterin in der Didaktik der Biologie der Freien Universität Berlin.

LITERATUR

Eckhardt, M./Germ, M./Großschedl, J.:

Der Flossenstrahl-Effekt – Natur als Lösungsquelle für technische Innovationen. In: Unterricht Biologie, Heft 332/2008, S. 12–16

Skiba, F./Spieler, M.:

Fische und Modell. In: Unterricht Biologie, Heft 315/2006, S. 42–43

Westheide, W./Rieger, R. (Hrsg.):

Spezielle Zoologie. Teil 2: Wirbel- oder Schädeltiere. Heidelberg 2010

UNTERRICHTSMATERIAL

Unterrichtsmaterial zu diesem Beitrag

finden Sie in der Heftmitte auf S. XIII–XVI: Schwimmen wie Fische im Wasser
Bau von Fischmodellen
Versuchsdurchführung
Notizen zum Versuch

Schwimmen wie Fische im Wasser

Eva und Jonas angeln in ihrer Freizeit. Eines Sonntags, als sie wieder mit ihren Angeln am Seeufer sitzen, stöhnt Jonas laut auf: „Ach, ich würde mich gerne auch so schnell und mühelos im Wasser bewegen können, wie diese Fische es tun. Wenn ich mich durch das Wasser bewege, dann komme ich immer nur langsam voran. Das ist so anstrengend.“ „Ja, so schwimmen können wie Fische im Wasser, das wäre toll!“, stimmt Eva ihm zu.

Eine Weile vergeht. Die Kinder schauen zu, wie die Fische im Wasser umherflitzen. Plötzlich ruft Eva: „Ich habe einen Fisch an der Angel. Schnell, Jonas, hilf mir, ihn hochzuziehen!“ Beide Kinder kurbeln kräftig an der Angel, bis der gefangene Fisch, der eine längliche Körperform hat, aus dem Wasser auftaucht. Während Eva die Angel hält, versucht Jonas den Fisch zu ergreifen, um ihn von der Angel zu lösen. Der Fisch zappelt jedoch wild umher und ist rutschig, sodass Jonas ihn nicht festhalten kann. Schließlich fällt der Fisch zurück ins Wasser. „Oh, das tut mir Leid, Eva. Jetzt ist er weg“, sagt Jonas. „Ach, nicht so schlimm, ich hätte ihn sowieso wieder frei gelassen. Ich wollte nur einmal von ganz nah seine glänzende Haut und seine Schuppen betrachten“, beruhigt ihn Eva. „Stimmt, die Schuppen habe ich auch deutlich erkennen können, und ich konnte fühlen, dass der Fisch von einer Schleimschicht überzogen war“, gab Jonas zu.



Notiert die Besonderheiten von Fischen, die Eva und Jonas während ihres Angelausfluges feststellen konnten!

Bau von Fischmodellen

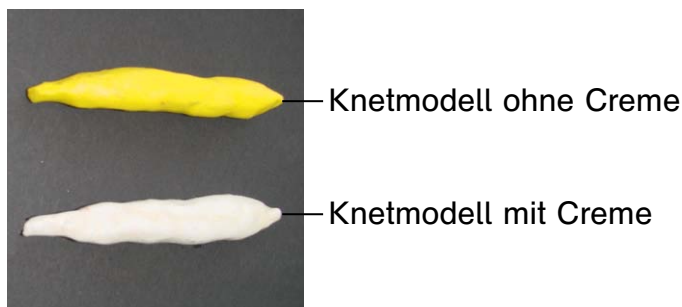
Ihr braucht:

- 30 g Knete für zwei Modelle
- Creme (z. B. Vaseline) für die Schleimschicht
- 2 Schnüre (jeweils 1 m)
- 2 Zahnstocher
- 2 kleine Gewichte (jeweils ca. 9 g)
- Wasser
- Blumenkasten (1 m)
- Tisch (Abstand zwischen Boden und Tischkante mindestens 1 m)
- Waage

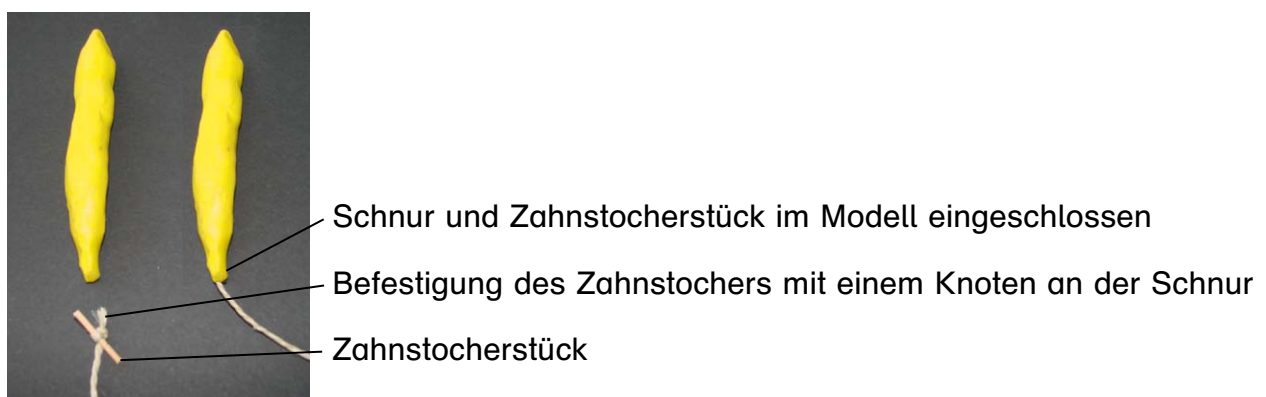
So geht es:

1. Formt mit Knete zwei Fischmodelle gleicher, z. B. länglicher Form.
2. Beschichtet ein Modell gleichmäßig mit Creme.

Wichtig: Das Modell mit Creme und das Modell ohne Creme müssen beide gleich schwer sein. Nutzt dazu die Waage.



3. Befestigt an beiden Modellen die Schnüre mit **gleich schweren Gewichten**. Ein kurzes Zahnstocherstück, an dem die Schnur befestigt wurde, verhindert das Lösen der Schnur aus der Knete.



4. Füllt den Blumenkasten zu $\frac{3}{4}$ mit Wasser.

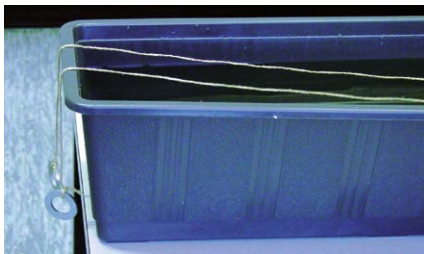
Versuchsdurchführung

Probedurchlauf

1. Legt beide Modelle nebeneinander in den Blumenkasten und haltet sie fest.



2. Spannt die Schnur durch das Wasser. Die Gewichte hängen dabei über der Tischkante herunter.



3. Gebt ein Startzeichen und notiert, welches der Modelle schneller ist.

Durchführung des Versuchs

Nach dem Probedurchlauf wird der Versuch zehnmal wiederholt. Notiert bei jedem Versuch, welches Modell schneller das Ende des Blumenkastens erreicht!

	Modell mit Creme	Modell ohne Creme
1. Versuch		
2. Versuch		
3. Versuch		
4. Versuch		
5. Versuch		
6. Versuch		
7. Versuch		
8. Versuch		
9. Versuch		
10. Versuch		

FOTOS: JULIANE GRÜNKORN, NICOLA HANAUER

Notizen zum Versuch

1. Die folgende Fragestellung wird untersucht:

2. Die folgende Hypothese ist möglich:

3. Notiere: Was müsst ihr bei diesem Versuch bedenken?

Überlege und bedenke, was du bereits über Fische und die Fischhaut weißt.
Bedenke auch, was du bereits über den Bau von Modellen weißt.

4. Führt den Versuch mit Modellen durch.

Notiert eure Beobachtungen/Ergebnisse. Wie oft war welches Modell schneller?

5. Überlegt und notiert eure Schlussfolgerung. Welche Bedeutung hat das Ergebnis?

Beitrag 6

Grünkorn, J. & Fleige, J. (2012). Bau und Funktion der Fischhaut. In J. Fleige, A. Seegers, A. Upmeyer zu Belzen & D. Krüger (Hrsg.), *Modellkompetenz im Biologieunterricht Klasse 7-10. Phänomene begreifbar machen – in 11 komplett ausgearbeiteten Unterrichtseinheiten* (S. 23-28). Donauwörth: Auer.



Ein Unterrichtskonzept von Juliane Grünkorn und Jennifer Fleige

Jahrgang	Klasse 7/8
Zeitumfang	90 Minuten
Unterrichtsreihe	Fische (Angepasstheit an den Lebensraum Wasser)
Fachinhalt	Die Haut der Fische ist u. a. mit Schuppen besetzt, die durch ihre Anordnung Schutz bieten und zugleich eine gute Beweglichkeit gewährleisten.
Kompetenzen MK	Die SuS erstellen verschiedene Fischschuppen-Modelle und nutzen diese sachgerecht, indem sie mit diesen Modellen Hypothesen über die Anordnung der Fischschuppen und deren Effekt auf die Eigenschaften der Fischhaut aufstellen, diese Hypothesen überprüfen und die Modelle ggf. ändern.
Methoden	Modellbau in Gruppenarbeit, Unterrichtsgespräch
Materialien	Kopien der Arbeitsblätter (AB 1 und 2), AB 2 als Folie, Film ¹ über schwimmenden Fisch (alternativ: Fisch im Aquarium), feste Pappe, Papier, Schere, schnell trocknender Klebstoff, echter Fisch (alternativ: Abbildung (z. B. Abb. 3) oder Fischhautpräparat)

➔ **Einstieg:** Zu Beginn der Stunde wird ein schwimmender Fisch gezeigt (z. B. Schuppenkarpfen im Aquarium, Film), um mit den SuS die Merkmale von Fischen zu wiederholen. Der Aspekt, dass Fische Schuppen haben, wird hervorgehoben. L: „Wir untersuchen, welchen Effekt die Anordnung der Fischschuppen auf die Eigenschaften der Fischhaut hat.“ Die Fragestellung wird auf **AB 1** notiert.

➔ **Erarbeitung:** In Kleingruppen entwickeln die SuS Vorüberlegungen zur Anordnung der Fischschuppen und notieren diese (**AB 1**, Aufgabe 1). Anschließend bauen sie auf dieser Grundlage ein Modell. Die zur Verfügung stehenden Materialien sind feste Pappe, Schere, schnell trocknender Kleber (z. B. Sekunden- oder Fotokleber) und ein weißes Blatt Papier. Die SuS schneiden Schuppen aus der Pappe aus, kleben diese auf das Papier und beschreiben zunächst ihr gebautes Modell (**AB 1**, Aufgabe 2 und 3).

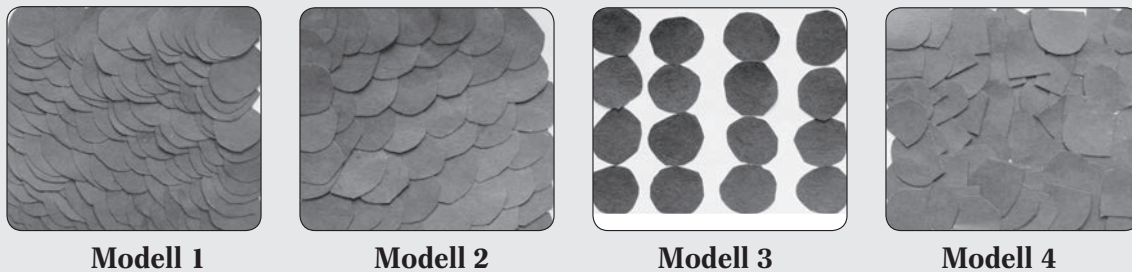


Abb. 1: Mögliche Schülermodelle

In einem weiteren Schritt sollen die SuS ihr Modell biegen, um in Gruppen über den Effekt der von ihnen gewählten Schuppenanordnung zu diskutieren und Schlussfolgerungen auf die Eigenschaften der Fischhaut ziehen zu können. Die am Modell abgeleiteten Voraussagen halten die SuS fest (**AB 1**, Aufgabe 4).

➔ **Sicherung:** Die Gruppen stellen im Plenum ihr gebautes Modell (**Abb. 1**) und ihre Schlussfolgerungen vor. Die verschiedenen Modellbeschreibungen werden an der Tafel festgehalten (**Abb. 2**: Beschreibung des Modells). Anschließend werden die gebauten Modelle miteinander verglichen und die unterschiedliche Auswirkung auf die Beweglichkeit und den Schutz der Fischhaut herausgearbeitet. Durch Anfassen und Biegen aller

MK

¹ z. B. <http://www.youtube.com/watch?v=VltAaH2kQRI&feature=related>



Modelle können Unterschiede erspürt werden. Die Ergebnisse werden ebenfalls an der Tafel notiert (**Abb. 2**: Beweglichkeit und Schutz). Die SuS leiten aus den Diskussionen über die verschiedenen Modelle die Anordnung der Fischeschuppen ab, die am wahrscheinlichsten ist. Wenn sich die SuS nicht für eine Schuppenanordnung entscheiden können, werden mehrere Schuppenanordnungen angenommen. In einem nächsten Schritt schlagen die SuS das weitere Vorgehen im Erkenntnisprozess vor: Ein genauere Blick auf die Fischhaut ist nötig. Hier sollte ein echter Fisch (z. B. Fischhautpräparat, tiefgekühlter Fisch) herangezogen werden (**Abb. 3**: Daten 2). Die gemeinsam abgeleitete Vermutung über die Anordnung der Fischeschuppen sowie die Vorüberlegungen der Gruppen werden mit der echten Fischhaut verglichen. So können einige Vermutungen über die Anordnung der Fischeschuppen falsifiziert und die dazugehörigen Modelle verworfen oder evtl. geändert werden (**Abb. 1**: Modell 1, 3 und 4). Modell 2 kann als vorläufig angenommen werden. Modelländerungen können bei ausreichender Zeit durchgeführt oder zumindest theoretisch besprochen und als Ergebnis an der Tafel (**Abb. 2**: Änderungen am Modell) und auf dem Arbeitsblatt festgehalten werden (**AB 1**, Aufgabe 5).

T

Ä

MK

→ **Reflexion:** Die SuS füllen in Partnerarbeit das Schema „Der Einsatz von Modellen zum Erkenntnisgewinn“ am Fischeschuppenbeispiel aus (**AB 2**). Im Plenum werden die Lösungen besprochen und L stellt gezielt Fragen über die Modelle (**Tab. 1**: Legende E – Ä). Die Ergebnisse sollen gesammelt werden (z. B. Folie von **AB 2**).

→ **Thematische Anknüpfungspunkte:** In weiteren Stunden können die gebauten Fischeschuppen-Modelle im Sinne einer „Kosten-Nutzen-Analyse“ bewertet werden. Die SuS wägen ab, bei welchen Schuppen-Modellen das Verhältnis zwischen „Nutzen“ (Schutz, Beweglichkeit) zu „Kosten“ (Material- und Energieverbrauch) optimal ist. Weiterhin kann beim Vergleich zwischen den Fischeschuppen-Modellen mit regelmäßiger und unregelmäßiger Schuppenanordnung über den Aspekt Strömungswiderstand gesprochen werden. In diesem Zusammenhang können auch andere Bestandteile der Fischhaut, wie z. B. die Schleimschicht, thematisiert werden.



Name:

Datum:

Folgende Fragestellung soll untersucht werden:

1. Notiere deine Vorüberlegungen zur Fischschuppenanordnung.

2. Baue mit den bereitgestellten Materialien ein Modell der Fischhaut. Schneide zunächst Pappschuppen aus. Die Schuppen sollten ungefähr 1,5 cm × 1,5 cm groß sein. Klebe diese dann auf das weiße Papier.

3. Beschreibe dein Fischschuppen-Modell.

4. Biege dein Fischschuppen-Modell. Beschreibe, welchen Effekt dein Fischschuppen-Modell auf die Eigenschaften der Fischhaut hätte.

5. Ergebnisse aller Gruppen: Diese Tabelle wird nach der Gruppenpräsentation ausgefüllt.

Beschreibung des Modells	Beweglichkeit	Schutz	Änderungen am Modell



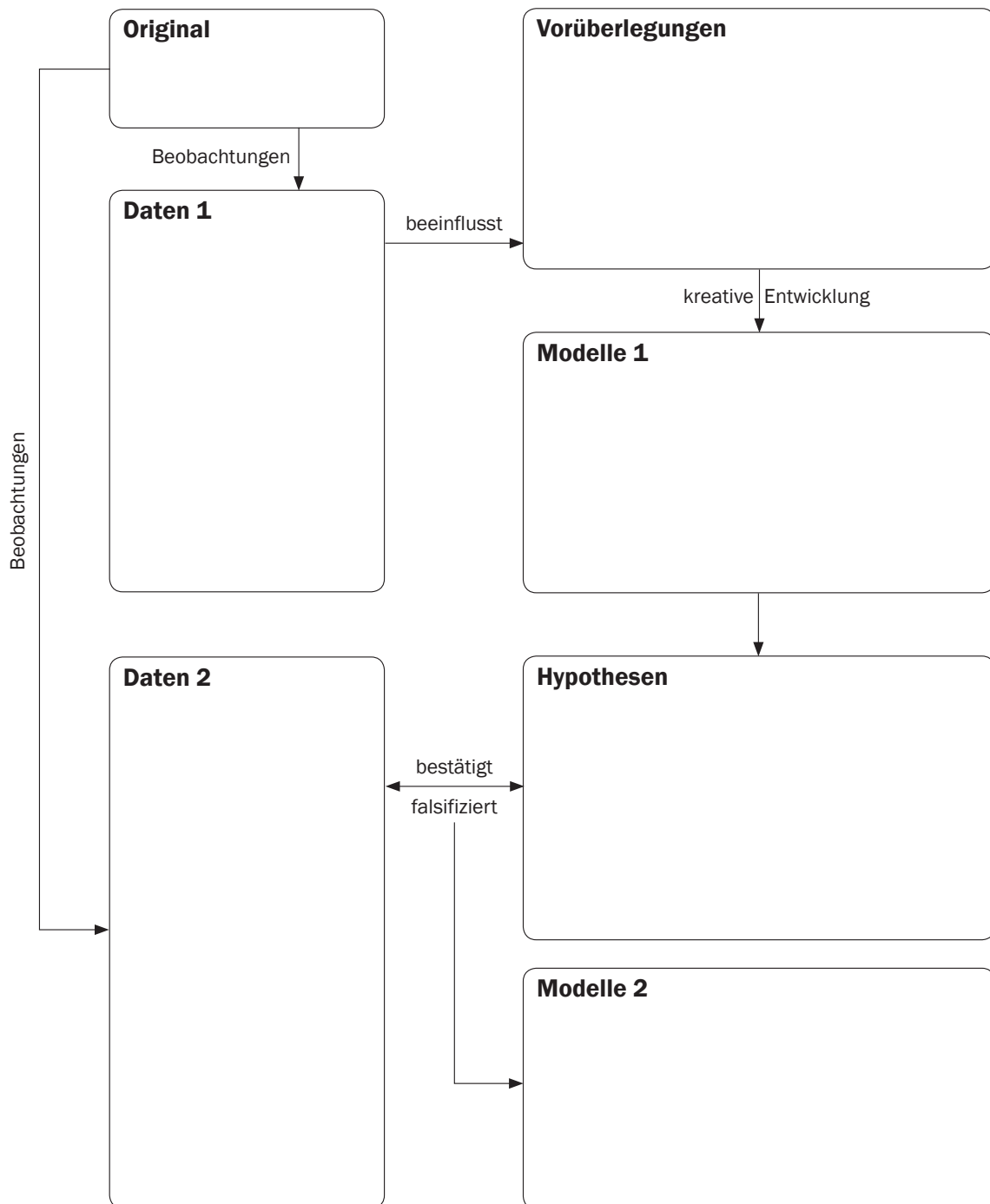
Name:

Datum:

Der Einsatz von Modellen zum Erkenntnisgewinn

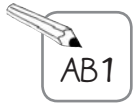
Fülle mit deinem Nachbarn das Schema zur Erkenntnisgewinnung mit Modellen aus.

Fragestellung:





→ Lösungsvorschläge:



Aufgabe 1–4: Individuelle Schülerlösungen

Aufgabe 5: z. B. Abb. 2

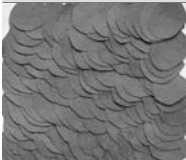
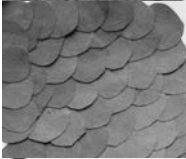
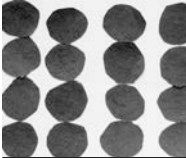
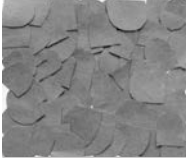
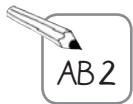
Auswirkung auf die Eigenschaften der Fischhaut				
Beschreibung des Modells		Beweglichkeit	Schutz	Änderungen am Modell
	sehr dicht übereinanderliegende Schuppen	-	+	muss geändert werden → Schuppen weiter auseinander
	dachziegelartig übereinanderliegende Schuppen	+	+	ist vorläufig gültig
	nebeneinanderliegende Schuppen	+	-	muss geändert werden → Schuppen dichter übereinander
	unregelmäßig übereinanderliegende Schuppen	+/-	+/- ¹	muss geändert werden → Schuppen regelmäßig übereinander

Abb. 2: Mögliches Tafelbild



Komplexität / Teilkompetenz	Niveau		
	Niveau I	Niveau II	Niveau III
Eigenschaften von Modellen	Begründe, inwieweit dein Modell so aussieht wie das Original.		
	Die Modelle sehen (fast) so aus wie eine echte Fischhaut.	Die Modelle zeigen vereinfacht bzw. im Wesentlichen, wie die echte Fischhaut aussieht.	Die Modelle stellen eine Vermutung über die mögliche Anordnung der Fischeschuppen dar.
Alternative Modelle	Begründe, warum es verschiedene Modelle gibt.		
	Es gibt verschiedene Modelle, weil sie unterschiedlich dargestellt (u. a. Farbe, Größe) sind.	Es gibt verschiedene Modelle, weil sie unterschiedliche Aspekte des Originals (Struktur und/oder Funktion, verschiedene Abschnitte der Fischhaut) berücksichtigen.	Es gibt verschiedene Modelle, weil diesen unterschiedliche Vermutungen über die Anordnung der Fischeschuppen zugrundeliegen.
Zweck von Modellen	Beschreibe, welchen Zweck dein Modell hat.		
	Die Modelle zeigen, wie die Fischhaut aussieht.	Die Modelle erklären, welche Zusammenhänge zwischen der Anordnung der Fischeschuppen und dem Schutz bzw. der Beweglichkeit der Fischhaut bestehen.	Die Modelle stellen Vermutungen über die Anordnung der Fischeschuppen dar und ermöglichen durch eine Anwendung der Modelle eine Vorrausage über die Eigenschaften (Schutz und Beweglichkeit) der Fischhaut.

¹ Durch eine unregelmäßige Anordnung der Schuppen können Stellen der Fischhaut unbedeckt sein, wodurch ein vollständiger Schutz der Fischhaut nicht gewährleistet wird.



Wissenschaftliches Arbeiten mit Modellen – Lösungen

Bau und Funktion der Fischhaut

Komplexität / Teilkompetenz	Niveau I	Niveau II	Niveau III
Testen von Modellen	Erkläre, wie man prüfen kann, ob dein Modell tauglich ist.		
	Wir überprüfen die Modelle, indem wir sie biegen und damit die Stabilität/Festigkeit des Materials prüfen.	Wir überprüfen die Modelle, indem wir sie mit einer echten Fischhaut vergleichen.	Wir überprüfen die Modelle, indem wir sie z. B. biegen und damit unsere Vermutungen über die Anordnung der Fischschuppen/ die Eigenschaften der Fischhaut überprüfen können.
Ändern von Modellen	Begründe, was dazu führen könnte, dass dein Modell verändert werden muss.		
	Die Modelle müssen verändert werden, wenn die Materialien nicht stabil genug sind.	Die Modelle müssen verändert werden, wenn neue Informationen über die echte Fischhaut vorliegen.	Die Modelle müssen verändert werden, wenn die dem Modell zugrundeliegenden Vermutungen widerlegt werden.

Tab. 1: Der Einsatz von Modellen zum Erkenntnisgewinn¹

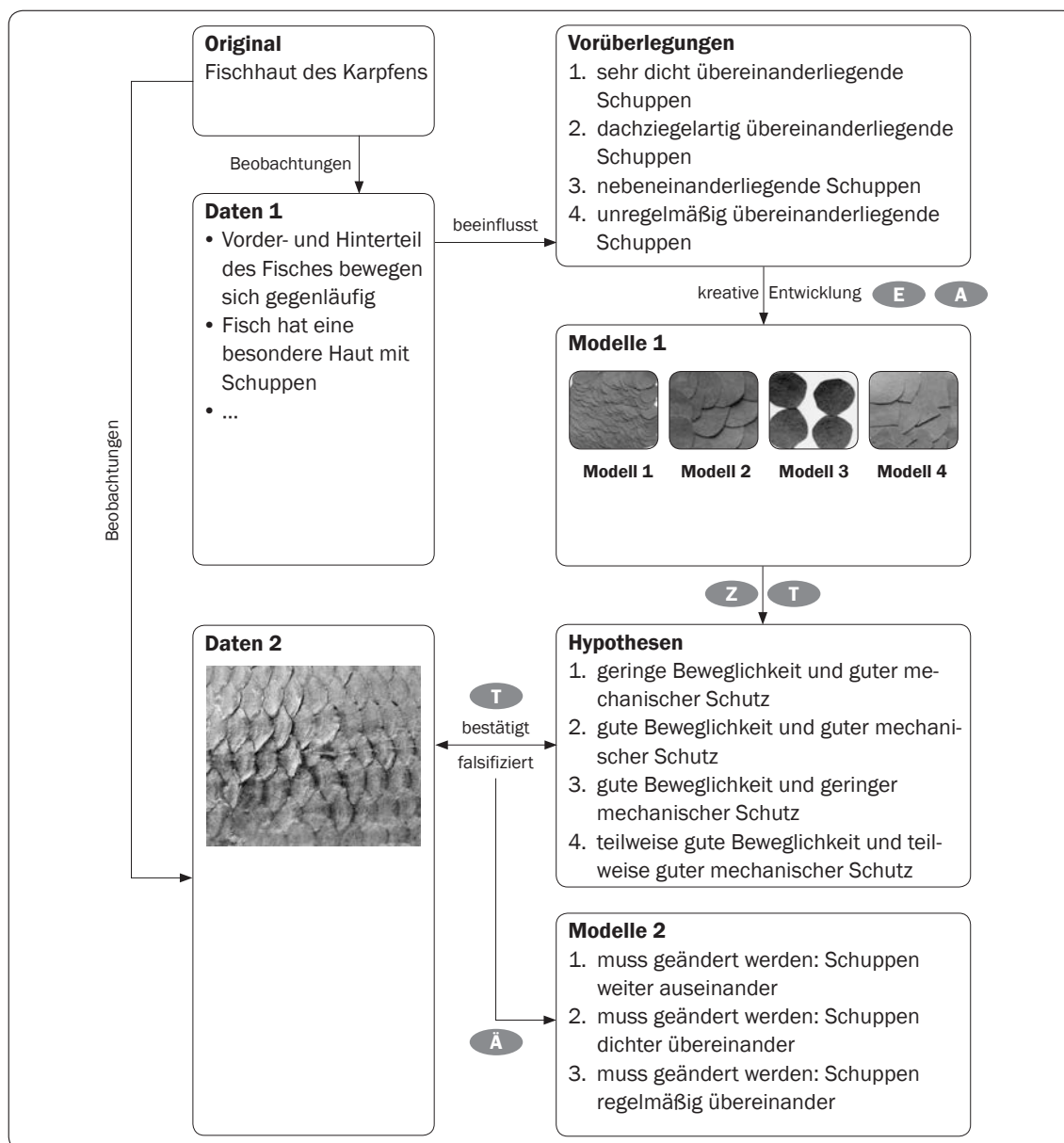


Abb. 3: Reflexionsschema

¹ Da es in bestimmten Unterrichtsphasen ebenfalls wichtig ist, über Niveau I und II zu reflektieren, werden bei diesem Beispiel Antwortmöglichkeiten auf allen drei Niveaus formuliert.

Danksagung

„Zwei Dinge sind zu unserer Arbeit nötig: Unermüdliche Ausdauer und die Bereitschaft, etwas, in das man viel Zeit und Arbeit gesteckt hat, wieder wegzwerfen.“ (Albert Einstein)

Dieses Werk ist erst durch die unterschiedlichste Unterstützung zahlreicher Personen entstanden, denen ich an dieser Stelle sehr gerne danken möchte.

Beginnen möchte ich mit den Schülerinnen und Schülern und Lehrkräften, die sich bereit erklärt haben, an meinen Studien teilzunehmen und damit die Grundlage für meine Dissertation geschaffen haben. Der Deutschen Forschungsgemeinschaft (DFG) sei für die finanzielle Unterstützung eines Teils dieses Promotionsprojekts gedankt.

Ich danke meinem Doktorvater, Prof. Dr. Dirk Krüger, dafür, dass er mir die Möglichkeit gegeben hat, mich diesem Thema widmen zu dürfen und mir damit den Weg in die empirische Bildungsforschung geebnet hat. Besonders dankbar bin ich für den kritischen Austausch und die gute und meiner Meinung nicht unwichtige Vorbereitung auf und Betreuung während nationaler und internationaler Tagungen.

In diesem Zusammenhang sei auch meiner Doktormutter, Prof. Dr. Annette Upmeier zu Belzen, sehr gedankt, die mich in meiner Promotionszeit gemeinsam mit meinem Doktorvater betreut und begleitet hat. Ihre Ruhe aber auch klare Sicht auf Dinge hat nicht unwesentlich dazu beigetragen, dass diese Dissertation entstanden ist.

Weiterhin sei den Kolleginnen und Kollegen der Arbeitsgruppe der Biologiedidaktik der Freien Universität (FU) Berlin für die Unterstützung besonders während der Erhebungsphasen gedankt. Auch die fachlichen Diskussionen mit den Doktorandinnen und Doktoranden der FU-Arbeitsgruppe und der Modellgruppe haben mir sehr weitergeholfen.

Für den Rat und die konstruktiven Hinweise von Dr. Alexander Naumann und Prof. Dr. Johannes Hartig hinsichtlich methodischer Fragen bedanke ich mich genauso wie für die Unterstützung von Prof. Dr. Eckhard Klieme, Beate Abrie und Nefise Özmen zur Endphase meiner Dissertation.

Dr. Cornelia Sander, Sarah Huch und Sarah Dannemann danke ich für die unzähligen konstruktiven Hinweise, Tipps und Ideen zu meiner Arbeit. Eine große

und sehr verlässliche Hilfe war in dieser Zeit Nicola Hanauer. Sie erledigte viele Aufgaben sehr autark und hat es mir dadurch möglich gemacht, dass ich mich auf das Wesentliche konzentrieren konnte. Christiane Patzke und Dr. Sandra Nitz danke ich für den konstruktiven Austausch im Hinblick auf meine Arbeit. Dr. Eva Terzer hat mich während meiner gesamten Promotionszeit begleitet. Ich danke ihr für die zahlreichen fruchtbaren und konstruktiven Gespräche und den damit verbundenen Gedankenaustausch. Euch allen sei sehr gedankt und um es mit **anderen Worten zu sagen: Ihr bekommt definitiv alle „ein Foto“.**

Schließlich möchte ich noch meinen Eltern, meiner Schwester, Diana Grünkorn, und besonders Hannsjörg Braun einen Dank für die Geduld und die liebevolle Unterstützung in dieser Zeit aussprechen.

Anhang

Anhang 1:	Kodierleitfäden.....	266
Anhang 2:	Testmanual der Teilstudien 1 und 2.....	280
Anhang 3:	Konstruktionsanleitungen für Aufgaben im offenen Antwortformat.....	283
Anhang 4:	Testmanual der Hauptstudie.....	288
Anhang 5:	Kennwerte der zweidimensionalen IRT-Skalierung.....	291
Anhang 6:	Publikationsliste.....	294

Anhang 1: Kodierleitfäden

Theoretische Grundlagen

Das Kompetenzmodell der Modellkompetenz (Upmeier zu Belzen & Krüger, 2010) basiert auf empirischen Studien (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003) sowie wissenschaftstheoretischer Literatur zu Modellen (Giere, 2004; Mahr, 2008). Es ist in fünf Teilkompetenzen – *Eigenschaften von Modellen*, *Alternative Modelle*, *Zweck von Modellen*, *Testen von Modellen* und *Ändern von Modellen* – strukturiert. Jede Teilkompetenz unterscheidet sich in drei Niveaus, die ein unterschiedliches Maß an Reflexions- und Abstraktionsvermögen ausdrücken.

Kompetenzmodells der Modellkompetenz (verändert nach Upmeier zu Belzen & Krüger, 2010)

Teilkompetenz	Komplexität		
	Niveau I	Niveau II	Niveau III
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	Modell als Modell von etwas durch neue Erkenntnisse oder zusätzliche Perspektiven revidieren	Modell für etwas aufgrund falsifizierter Hypothesen revidieren

Allgemeiner Hinweis zum Kodieren

1. Im Transkript bedeuten eckige Klammern [], dass Wörter vom Dateneingebener hinzugefügt wurden, um den Satz besser verstehen zu können.
2. Im Transkript bedeuten eckige Klammern plus das darin durchgestrichene Wort [~~Wort~~], dass Wörter vom Dateneingebener gestrichen wurden, um den Satz besser verstehen zu können.

Kodierregeln

1. Die Aussagen werden den vorgegebenen Codes zugeordnet.
2. Bei Unsicherheiten bezüglich der Zuordnung oder diskussionswürdigen Auffälligkeiten werden die entsprechenden Aussagen mit einem erklärenden Memo versehen.
3. Bei der Kodierung wird so vorgegangen, dass pro Aussage minimal ganze Sätze (Kodiereinheit) oder – falls dies für die Nachvollziehbarkeit der Zuordnung notwendig ist – maximal ganze Absätze (Kontexteinheit) dem Code zugeordnet werden. Es werden keine Satzteile, sondern immer ganze Sätze codiert.

Beispiel für falsches Zuordnen

Aussage	Zugeordnete Aussage zum Code
Die Modelle sind Kopien vom Original.	Kopien falsch

Beispiel für richtiges Zuordnen

Aussage	Zugeordnete Aussage zum Code
Die Modelle sind Kopien vom Original.	Die Modelle sind Kopien vom Original. richtig

4. Einzelne Sätze oder ganze Aussagen können auch **mehreren Codes zugeordnet** werden. Wenn das der Fall ist, sollte in einem Memo erklärt werden, welcher Teil des Satzes welchem Code zugeordnet wurde.

Beispiel für die Zuordnung einer Aussage zu mehreren Codes

Aussage	Zugeordnete Aussage zum Code 1 (Kopie)	Zugeordnete Aussage zum Code 2 (Fokussierte Darstellung)
Das gezeigte Modell kann eine Kopie oder eine fokussierte Darstellung sein.	Das gezeigte Modell kann eine Kopie oder eine fokussierte Darstellung sein.	Das gezeigte Modell kann eine Kopie oder eine fokussierte Darstellung sein.

5. Wenn eine Schülerin bzw. ein Schüler diese Aufgabe nicht beantwortet hat, wurde dies mit „9999“ gekennzeichnet. Diese Kennzeichnung wird dem Code „keine Antwort“ zugeordnet.
6. Wenn Aussagen nicht den Niveaus I, II oder III zuzuordnen sind, werden diese dem Code „Sonstiges“ zugeteilt. Hier gegebenenfalls Unterkodes generieren (z. B. keine Ideen, keine Korrespondenz zur Frageintention)
7. Wenn Aussagen zwar einem Niveau zugeordnet werden können, allerdings kein passender Code vorhanden ist, sollte ein neuer Code hinzugefügt, benannt und beschrieben werden.

Kodierschema der Teilkompetenz „Eigenschaften von Modellen“ Teil I

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	Keine Antwort <ul style="list-style-type: none"> Lernende haben die Aufgabe erhalten, diese aber nicht beantwortet. 	9999
	Sonstiges <ul style="list-style-type: none"> Lernende geben eine Antwort, die nicht zugeordnet werden kann, weil sie nicht mit der Frageintention korrespondiert. 	<i>Der Tyrannosaurus rex ist ein sehr großer Saurier. Sie waren gefährlich.</i>
	Modell als Mittel der Zugänglichkeit <ul style="list-style-type: none"> dient dazu, Dinge für einen bestimmten Adressaten zugänglich bzw. direkt erfahrbar zu machen 	<i>Es hat den Zweck, Dinge darzustellen, die nicht für alle Menschen zugänglich sind.</i>
	Modell als Mittel der Verständlichkeit/Kommunizierbarkeit <ul style="list-style-type: none"> verbessert das Verständnis und die Kommunizierbarkeit /erleichtert die Vermittlung 	<i>Es sieht gleich aus, aber ist vielleicht nur vergrößert, weil man den Zweck dann vielleicht besser erklären kann.</i>
	Modell als Kopie <ul style="list-style-type: none"> gleicht dem Original vergrößerte/verkleinerte Kopie des Originals als maßstabsgetreues Duplikat des Originals akzeptiert, weil ein großes Vertrauen gegenüber der Wissenschaft, den wissenschaftlichen Methoden oder den Forschenden besteht 	<i>Ich glaube, dass der Tyrannosaurus rex so aussah wie das Modell.</i> <i>Meiner Meinung nach ist das Modell einer Biomembran eine vergrößerte Kopie einer Biomembran.</i> <i>Ich glaube, dass die Neandertaler auch so aussahen wie das Modell, weil bestimmt viele Biologen an diesem Modell gearbeitet haben. Diese Leute kennen sich bestimmt mit sowas aus, also glaube ich schon, dass er so aussah.</i>
	Modell mit großer Ähnlichkeit <ul style="list-style-type: none"> ähnelt dem Original nahezu originalgetreues Duplikat des Originals aus Unzufriedenheit mit dem Modellierungsprozess 	<i>Das Modell der Biomembran ähnelt der echten Biomembran sehr. Beide haben einen Rand, der alles zusammen hält und beide haben in der Mitte ein Gewebe.</i> <i>Das Modell ähnelt dem Neandertaler. Nur die Stelle an der die Augenbrauen von dem Modell sind, müsste etwas nach vorne gedrückt werden, da es im Knochenfund so ist. Das müsste so sein, weil das Modell sonst nicht richtig wäre.</i>

Kodierschema der Teilkompetenz „Eigenschaften von Modellen“ Teil II

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
I	<p>Modell entspricht (nicht) subjektiver Vorstellung vom Original</p> <ul style="list-style-type: none"> • vergleichen und beurteilen das Modell nach eigenem Vorwissen, persönlichen Erfahrungen oder subjektiven Vorstellungen über das Original 	<p><i>Man kann sich gut vorstellen, dass das Modell eines Neandertalers so aussieht, da Neandertaler meiner Vorstellung nach so aussehen.</i></p> <p><i>Ich glaube nicht, dass das Modell richtig ist. Ich glaube, dass der echte Neandertaler mehr wie ein Affe ausgesehen hat. So stelle ich ihn mir nun mal vor.</i></p>
	<p>Modell ist in Teilen eine Kopie</p> <ul style="list-style-type: none"> • gleicht nur in bestimmten Merkmalen dem Original, andere Merkmale können bei geringer Information/Kennntnis über das Original nicht beurteilt werden 	<p><i>Der Knochenfund und das Modell haben die gleiche Kopfform. [...] Ob die Haare früher auch so waren, weiß man nicht. Auch über die Augen kann man nichts aussagen. Im Großen und Ganzen kann man immer nur etwas über die Form sagen. Farbe und so bleibt „unbekannt“.</i></p>
II	<p>Modell als eine mögliche Variante</p> <ul style="list-style-type: none"> • gleicht möglicherweise (nicht) dem Original, abstrakte Aussage zu übereinstimmenden Merkmalen • eine denkbare Version unter vielen, jedoch wenig begründet 	<p><i>Das Modell ist im Bezug auf seine Form vergleichbar. Trotzdem kann man nicht davon ausgehen, dass der Neandertaler wirklich so ausgesehen hat.</i></p> <p><i>Ja, ich glaube schon, dass sie [die Biomembran] so aussehen könnte, aber sie könnte auch so aussehen [gemaltes Bild der Schülerin].</i></p>
	<p>Modell als fokussierte Darstellung</p> <ul style="list-style-type: none"> • fokussiert auf einen Ausschnitt des Originals, hebt bestimmte Merkmale/Eigenschaften hervor 	<p><i>Ich denke, dass das Modell nur das Wesentliche einer echten Biomembran zeigt. Die wesentlichen Merkmale, Strukturen und Farben werden hier gezeigt.</i></p>
III	<p>Modell als hypothetische Darstellung</p> <ul style="list-style-type: none"> • stellt begründet eine Hypothese über das Original dar, mögliche Ähnlichkeit zwischen Original und Modell wird diskutiert 	<p><i>Man kann nicht genau wissen, wie der damals lebende Tyrannosaurus rex aussah. Die Wissenschaftler/innen können bloß Vermutungen anstellen, wie er aussah. Sie analysieren das Knochen skelett, und können dann berechnen, wie sein Körper aufgebaut sein könnte.</i></p>

Kodierschema der Teilkompetenz „Alternative Modelle“ Teil I

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	<p>Keine Antwort</p> <ul style="list-style-type: none"> Lernende haben die Aufgabe erhalten, diese aber nicht beantwortet. 	9999
	<p>Sonstiges</p> <ul style="list-style-type: none"> Lernende geben eine Antwort, die nicht zugeordnet werden kann, weil sie nicht mit der Frageintention korrespondiert. Lernende können sich nicht erklären, warum es verschiedene Modelle zu einem Original gibt. 	<p><i>Weil den Biologen sonst langweilig wäre.</i></p> <p><i>Keine Ahnung, warum es drei Modelle gibt. Ich kann mir das nicht erklären.</i></p>
	<p>Modell als Mittel der Zugänglichkeit</p> <ul style="list-style-type: none"> dient dazu, Dinge für einen bestimmten Adressaten zugänglich bzw. direkt erfahrbar zu machen 	<i>Es hat den Zweck, Dinge darzustellen, die nicht für alle Menschen zugänglich sind.</i>
	<p>Modell als Mittel der Verständlichkeit/Kommunizierbarkeit</p> <ul style="list-style-type: none"> verbessert das Verständnis und die Kommunizierbarkeit /erleichtert die Vermittlung 	<i>Ich denke, es gibt drei Modelle, da jeder Mensch es bei einem anderen Modell vielleicht besser versteht. Nicht jeder Mensch denkt ja gleich.</i>
	<p>Modelle sind gleich</p> <ul style="list-style-type: none"> Modelle sind alle gleich/keine Beschreibung von Unterschieden zwischen den Modellen 	<i>Alle drei zeigen im Grunde dasselbe, ich weiß nicht, warum es drei verschiedene Modelle gibt. Das macht keinen Sinn.</i>
Basal	<p>Verschiedene Modelle zu unterschiedlichen Originalen</p> <ul style="list-style-type: none"> jedes Modell repräsentiert ein anderes Original 	<i>Es könnte aber auch sein, dass man drei verschiedene Modelle macht, um die Biomembranen von verschiedenen Lebewesen, z. B. Mensch, Vogel und einer Kuh zu zeigen.</i>

Kodierschema der Teilkompetenz „Alternative Modelle“ Teil II

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	Nur ein endgültiges und richtiges Modell	
Basal	<ul style="list-style-type: none"> nur eines der alternativen Modelle ist korrekt, die anderen sind falsch nur eines der Modelle ist das beste/richtige/finale Modell; die gezeigten Modelle sind nicht zeitgleich gültig 	<p><i>Vielleicht sind nicht alle drei richtig. Vielleicht weisen Modelle Fehler oder Lücken auf und nur eins ist davon richtig. Warum sollte es sonst drei Modelle geben?</i></p> <p><i>Ich denke, dass die Modelle alte Modelle sind. Zu dieser Zeit hatte man vielleicht nicht die Informationen, die man heute hat. Eines davon ist aber das endgültige Modell, das auch stimmt. Ich habe es in meinem Biologiebuch gesehen.</i></p>
	Unterschiedliche Modellobjekteigenschaften	
I	<ul style="list-style-type: none"> verschiedene Darstellungsweisen (u. a. 2-D/3-D, verschiedene Farben) verschiedene Modellmerkmale (u. a. beweglich/unbeweglich, weich/hart) unterschiedliche Konstruktionsmöglichkeiten (u. a. dünne/dicke Materialien, getrennte/nicht getrennte Bausteine) unterschiedliche Übersichtlichkeit, Abstraktheit 	<p><i>Ich denke, dass das Modell A dazu dient die Biomembran grafisch darzustellen... Die Modelle B und C sind bildliche Modelle...</i></p> <p><i>Modell B und C werden unterschieden durch ihre Beschaffenheit. Während Modell B eher steif ist, ist Modell C sehr flexibel.</i></p> <p><i>Bei Modell A sind die verschiedenen Bausteine gut zu erkennen. [...] Für mich sind die Stäbchen auf Modell C dünner [als] bei B.</i></p> <p><i>Man kann sie [die Modelle] einfach, spezialisiert, kompliziert, groß, klein, realistisch, unrealistisch machen.</i></p>
	Unterschiedliche inhaltliche Schwerpunkte	
II	<ul style="list-style-type: none"> Komplexität des Originals ermöglicht unterschiedliche Perspektiven bzw. Foci auf das Original (u. a. innen/außen, Längs- /Querschnitt, Struktur/Funktion, verschiedene Ausschnitte bzw. Zustände des Originals) 	<p><i>Da jedes dieser Modelle etwas anders verdeutlicht und hervorhebt, gibt es unterschiedliche Modelle. Im Modell A geht man auf die verschiedenen Bausteine und die Struktur ein, im Modell B/C geht es eher um den Aufbau einer Biomembran.</i></p>
	Unterschiedliche Annahmen	
III	<ul style="list-style-type: none"> verschiedene Annahmen/Ideen über das Original/unterschiedliche Modelle sind alle zeitgleich gültig unterschiedliche Interpretationen der Daten 	<p><i>[...] Die einzelnen Personen haben aus ihren Beobachtungen unterschiedlich geschlussfolgert und deswegen gibt es verschiedene Modelle zu dieser Biomembran.</i></p>

Kodierschema der Teilkompetenz „Alternative Modelle“ Teil III

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
III	Unterschiedliche Annahmen mit Anwendungsperspektive <ul style="list-style-type: none">• verschiedene Annahmen über das Original werden mit wissenschaftlichem Nutzen genannt (u. a. Diskussionsgrundlage, Vergleich verschiedener Annahmen, Überprüfung der Annahmen mit den Modellen)	<i>Verschiedene Wissenschaftler haben [...] verschiedene Vorstellungen zu der Biomembran. [...]. Die Modelle dienen der Diskussion und dem Austausch mit anderen über den Aufbau der Biomembran.</i>

Kodierschema der Teilkompetenz „Zweck von Modellen“ Teil I

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	<p>Keine Antwort</p> <ul style="list-style-type: none"> Lernende haben die Aufgabe erhalten, diese aber nicht beantwortet. 	9999
	<p>Sonstiges</p> <ul style="list-style-type: none"> Lernende geben eine Antwort, die nicht zugeordnet werden kann, weil sie nicht mit der Frageintention korrespondiert. Lernende haben keine Idee, welchen Zweck das gezeigte Modell haben kann. 	<p><i>Damit es nicht sinnlos rumliegt... (F 338)</i></p> <p><i>Keine Ahnung, ich weiß keinen Zweck. (F 150)</i></p>
	<p>Modell als Mittel der Verständlichkeit/Kommunizierbarkeit</p> <ul style="list-style-type: none"> verbessert das Verständnis und die Kommunizierbarkeit/ erleichtert die Vermittlung 	<p><i>Es soll darstellen, wie ein Wald aussieht bzw. ist ein Hilfsmittel der Veranschaulichung, um es sich besser vorzustellen.</i></p> <p><i>Das Modell könnte für den Biologieunterricht genutzt werden, um zu zeigen, wie sich verschiedene Pflanzen zusammen verhalten.</i></p>
	<p>Modell als Mittel der Zugänglichkeit</p> <ul style="list-style-type: none"> dient dazu, Dinge für einen bestimmten Adressaten zugänglich bzw. direkt erfahrbar zu machen 	<i>Modelle können für Leute erfunden werden, die so etwas im realen Leben nicht kennen und durch Modelle können sie so ungefähr sehen, wie das ist.</i>
	<p>Modell als Spielzeug</p> <ul style="list-style-type: none"> dient dem Vergnügen, dem Interesse bzw. der Freude 	<i>Es gibt Kinder, die daran vielleicht interessiert sind und Spaß damit haben, deswegen wurde dieses Modell entwickelt.</i>
	<p>Modell zu Dekorationszwecken</p> <ul style="list-style-type: none"> dient der Verschönerung eines Raums/Gebäudes trägt zum Wohlbefinden von Personen und Lebewesen bei 	<p><i>Es kann sein, dass das Wald-Modell für zu Hause als Dekoration einen guten Zweck hat.</i></p> <p><i>Manche Leute beruhigt es [das Modell] auch einfach, wenn sie nicht gerade im Urlaub sind und einfach wegräumen können.</i></p>
	<p>Modell als Ersatzobjekt</p> <ul style="list-style-type: none"> dient als Prothese/Ersatzorgan 	<i>Der Zweck könnte sein, dass man ihn als Ersatzorgan entwickelt hat. Falls jemand einen defekten Darm besitzt, kann man ihn durch einen künstlich hergestellten voll funktionierenden Darm austauschen.</i>
	<p>Modell als Bauplan</p> <ul style="list-style-type: none"> dient als Grundriss im kleinen Maßstab, um das zukünftige Objekt darzustellen 	<i>Der Zweck des Modells ist, dass es eigentlich wie eine Landkarte benutzt wird. (...). Es dient sozusagen als Orientierung.</i>

Kodierschema der Teilkompetenz „Zweck von Modellen“ Teil II

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	Modell zu Werbezwecken <ul style="list-style-type: none"> dient als Werbemittel zum Erreichen von politischen, sozialen und/oder ökologischen Zielen 	<i>Damit wollen sie die Welt wecken, dass sie den Wald mehr bewässern.</i>
I	Modell zum Darstellen eines Sachverhaltes <ul style="list-style-type: none"> stellt Sachverhalte dar 	<i>Das Modell zeigt die verschiedenen Pflanzen, die [in] einem Wald vorkommen.</i>
	Modell zum Erkennen von Zusammenhängen <ul style="list-style-type: none"> beschreibt Zusammenhänge zwischen verschiedenen Aspekten im Original und dient dazu, bekannte Tatsachen nachzuvollziehen 	<i>Das Modell stellt dar, dass sie gucken kann, wie die Blätter und Blüten sich entwickeln und sich verbreiten.</i>
II	Modell zum Erklären von Zusammenhängen <ul style="list-style-type: none"> beschreibt und erklärt Zusammenhänge zwischen verschiedenen Aspekten im Original und dient dazu, bekannte Tatsachen nachzuvollziehen 	<i>Den Aufbau eines Waldes kann man daran auch erklären. Pflanzen (also auch Wälder) brauchen Erde, damit sie ihre Wurzeln ausschlagen können. Ebenfalls ist Wasser nötig, damit die Pflanzen wachsen.</i>
	Modell zum Überprüfen von abstrakten Ideen <ul style="list-style-type: none"> dient als Instrument zur Überprüfung von Hypothesen über das Original; allgemeine Ideen werden genannt 	<i>Vielleicht kann man gewisse Tests durchführen und so die Wirkung von bestimmten Dingen überprüfen.</i>
III	Modell zum Überprüfen von konkreten Ideen <ul style="list-style-type: none"> dient neben der Überprüfung von Hypothesen über das Original auch dazu, Schlüsse über das Original zu ziehen; konkrete Ideen werden genannt dient dazu, Erkenntnisse über das Original auf andere Phänomene zu übertragen 	<i>Zudem könnte man testen, welche Pflanze mit welchen Erdtypen am schnellsten und besten wächst und man könnte dies vergleichen. Möglicherweise könnte eine bestimmte Pflanze so viele Nährstoffe aus der Erde ziehen, dass einer anderen dann weniger zur Verfügung stehen. Wenn das Modell dies beweisen würde, wüsste man, dass man diese Pflanzentypen besser nicht auf engem Raum zusammen pflanzt.</i> <i>So ein Modell ist wahrscheinlich dazu da, um zu überprüfen, ob sich Pflanzen „vermehren“ können. Dies braucht man wiederum, um etwas über den Wald zu sagen und den Bestand dann in Zahlen oder in Diagrammen festzuhalten. Vielleicht kann man das dann auf andere Ökosysteme wie das Meer oder so beziehen.</i>

Kodierschema der Teilkompetenz „Testen von Modellen“ Teil I

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	Keine Antwort	
	<ul style="list-style-type: none"> Lernende haben die Aufgabe erhalten, diese aber nicht beantwortet. 	9999
	Sonstiges	
	<ul style="list-style-type: none"> Lernende geben eine Antwort, die nicht zugeordnet werden kann, weil sie nicht mit der Frageintention korrespondiert. Lernende nennen keine Ideen für ein Testen bzw. Überprüfen des Modells. 	<p><i>Ja, das Modell der Kleinlibelle ist tauglich und erfüllt seinen Zweck.</i></p> <p><i>Ich wüsste nicht, wie man es überprüfen könnte.</i></p>
	Modell als Mittel der Zugänglichkeit	
	<ul style="list-style-type: none"> dient dazu, Dinge für einen bestimmten Adressaten zugänglich bzw. direkt erfahrbar zu machen 	<i>Es hat den Zweck, Dinge darzustellen, die nicht für alle Menschen zugänglich sind.</i>
	Modell als Mittel der Verständlichkeit/Kommunizierbarkeit	
	<ul style="list-style-type: none"> verbessert das Verständnis und die Kommunizierbarkeit /erleichtert die Vermittlung 	<i>Das Modell der Libelle erfüllt den Zweck, wenn man bzw. Schüler damit das Verhalten des Tieres vielleicht besser verstehen können.</i>
	Keine Testung des Modells	
Basal	<ul style="list-style-type: none"> Ablehnung einer Testung des Modells an sich bzw. mit dem Modell 	<i>Warum soll man denn dieses Käfermodell testen??? Meiner Meinung nach ist dies unnötig.</i>
	Überprüfung des Materials	
	<ul style="list-style-type: none"> Überprüfung der Widerstandsfähigkeit des Materials (u. a. Beweglichkeit, Stabilität, Elastizität, Gewicht) 	<p><i>... Wenn es die Flügel aushalten, ist es tauglich...</i></p> <p><i>Das Modell könnte zu schwer sein, daher denke ich, die Flügel des Modells können die Last des nachgebauten Körpers nicht tragen.</i></p>
I	Überprüfung der Grundvoraussetzungen	
	<ul style="list-style-type: none"> Nennung grundlegender Anforderungen an das jeweilige Modell 	<p><i>Als erstes sollte das Modell doch auf jeden Fall fliegen können. Sonst ist das ganze Modell meiner Meinung nach nicht so gut...</i></p> <p><i>Man sollte es mit einer Fernbedienung steuern und es einfach mal ausprobieren.</i></p> <p><i>Man kann probieren, ob es fliegen kann.</i></p>

Kodierschema der Teilkompetenz „Testen von Modellen“ Teil II

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	Vergleich von Original und Modell <ul style="list-style-type: none"> Vergleich der Eigenschaften (Struktur und/oder Funktion) des Originals mit denen des Modells 	<i>Man kann es überprüfen, indem man das Modell mit einer echten Kleinlibelle vergleicht.</i>
II	Vergleich und Passung von Original und Modell <ul style="list-style-type: none"> zusätzlich zum Vergleich der Eigenschaften eine Beschreibung notwendiger Übereinstimmungen (Passung) von Modell und Original; Nennung von Kriterien für ein gutes Modell 	<i>Wenn das Modell so fliegen kann und so aussieht wie eine echte Libelle, dann erfüllt es seinen Zweck. Wenn nicht, dann muss man das anpassen.</i>
	Überprüfung von Hypothesen <ul style="list-style-type: none"> Überprüfung von Hypothesen über das Original mit dem Modell und Nennung allgemeiner Ideen für Untersuchungen 	<i>Dieses Modell könnte etwas simulieren. Darüber bekommt man dann Informationen über das Phänomen. Also testet man Auswirkungen von etwas auf etwas sozusagen die Vermutung, die man hat.</i>
III	Überprüfung von Hypothesen mit Forschungsdesign <ul style="list-style-type: none"> Beschreibung einer konkreten Anwendung des Modells (Forschungsdesign) zur Überprüfung einer Hypothese über das Original 	<i>Man müsste versuchen, Videos von Originalflugmanövern zu bekommen und dann diese mit dem Modell im Windkanal versuchen nachzuahmen und zu vergleichen, um dann zu sehen, ob sich das Modell so verhält wie das Original. Wenn dies der Fall ist, müsste man die Umwelteinflüsse im Windkanal verändern und somit Dinge herausfinden, die die Libelle zum Fliegen braucht. Dadurch erhält man immer neue Erkenntnisse über die Libelle.</i>

Kodierschema der Teilkompetenz „Ändern von Modellen“ Teil I

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	Keine Antwort	
	<ul style="list-style-type: none"> Lernende haben die Aufgabe erhalten, diese aber nicht beantwortet. 	9999
	Sonstiges	
	<ul style="list-style-type: none"> Lernende geben eine Antwort, die nicht zugeordnet werden kann, weil sie nicht mit der Frageintention korrespondiert. Lernende äußern explizit, dass sie keine Ideen für ein Ändern von Modellen haben. 	<p><i>Das Modellende sieht so aus, wie ein Boot, das die Libelle bald ziehen wird, finde ich einfach merkwürdig.</i></p> <p><i>Kein Plan, wie sich da etwas ändern sollte. Ich habe keine Idee.</i></p>
	Ändern des Modellzustandes	
	<ul style="list-style-type: none"> Zustandsänderung des Modells bei der Nutzung/Handhabung des Modells (u. a. Bewegungen des Modells) 	<i>Dieses Modell könnte verändert werden, da die Libellen im Flug viel machen kann. Dadurch verändert sie sich mit, weil sie sich bewegt.</i>
	Modell als Mittel der Zugänglichkeit	
	<ul style="list-style-type: none"> dient dazu, Dinge für einen bestimmten Adressaten zugänglich bzw. direkt erfahrbar zu machen 	<i>Es hat den Zweck, Dinge darzustellen, die nicht für alle Menschen zugänglich sind.</i>
	Modell als Mittel der Verständlichkeit/Kommunizierbarkeit	
	<ul style="list-style-type: none"> verbessert das Verständnis und die Kommunizierbarkeit /erleichtert die Vermittlung 	<i>Dieses Modell könnte verändert werden, weil vielleicht irgendetwas nicht ganz klar zum Ausdruck kommt, etwas, was die Schüler dann nicht so gut verstehen und das dann nicht so gut vermittelt werden kann.</i>

Kodierschema der Teilkompetenz „Ändern von Modellen“ Teil II

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	<p>Kein Anlass für eine Änderung</p> <ul style="list-style-type: none"> Ablehnung eine Änderung des Modells 	<p><i>Ich finde, dass man das Modell nicht verändern sollte. Warum sollte das Modell verändert werden? Das verstehe ich nicht.</i></p>
Basal	<p>Ändern zur Darstellung unterschiedlicher Originale</p> <ul style="list-style-type: none"> Herstellung verschiedener Modelle für unterschiedliche Originale; jedes Original wird durch ein Modell repräsentiert 	<p><i>Weil Libellen nicht alle gleich sind und weil man für verschiedene Libellen Modelle machen kann. Deswegen muss man das Modell verändern.</i></p>
	<p>Ändern zur Verbesserung des Modellobjekts</p> <ul style="list-style-type: none"> Optimierung der Funktionsfähigkeit/der Ästhetik des Modellobjekts Optimierung der Technik zur Modellherstellung 	<p><i>Der Grund zum Ändern von Modellen ist meistens nur, dass man die meisten Modelle noch verbessern kann in ihren Bewegungen und Funktionen.</i></p> <p><i>Das Modell der Libelle wird verändert werden müssen, da das Modell einen Motor und spezielle Gleitflügel hat und zusätzlich auch verschiedene Drähte und Kabel benötigt.</i></p>
I	<p>Ändern bei Fehlern im Modellobjekt</p> <ul style="list-style-type: none"> grundsätzliche Überlegungen zur Behebung von Fehlern am Modell Verweis auf konkrete fehlerhafte Eigenschaften des Modells (z. B. Materialmängel) 	<p><i>Ich denke, weil man immer wieder Fehler entdeckt und diese verbessert werden müssen.</i></p> <p><i>Vielleicht müsste man die Flügel aus härteren Materialien herstellen, da sie sonst dem Druck beim Fliegen nicht standhalten.</i></p>
	<p>Ändern bei Nichterfüllung der Grundvoraussetzungen</p> <ul style="list-style-type: none"> Überprüfung der grundlegenden Anforderungen an das jeweilige Modell und ggf. Behebung des Mangels 	<p><i>Ich müsste das Modell in der Realität vor mir haben, es anfassen können und z.B. schauen, ob es fliegt, erst dann könnte ich beurteilen, ob man an dem Modell etwas verändern müsste.</i></p> <p><i>Wenn das Modell zum Fliegen ist, und es das nicht macht, müssen die Wissenschaftler gründlich dran arbeiten.</i></p>

Kodierschema der Teilkompetenz „Ändern von Modellen“ Teil III

Niveau	Kategorienname mit Erläuterungen	Ankerbeispiele
	Ändern bei mangelnder Passung mit dem Original <ul style="list-style-type: none"> Optimierung der mangelnden Passung des Modells (Struktur und/oder Funktion) mit dem Original mit Blick auf notwendige Übereinstimmungen von Original und Modell 	<p><i>Wir haben beim hinteren Teil des Modells eine Art Tragfläche, die aber bei der Libelle nicht existiert. Das heißt, das Heck muss weg...</i></p> <p><i>Die Kleinlibelle steuert alles durch ihre Flügel am vorderen Teil ihres Rumpfes. Sie besitzt keine Heckstabilisatoren wie das Modell. ...</i></p>
II	Ändern bei neuen Erkenntnissen über das Original <ul style="list-style-type: none"> Integration neuer Erkenntnisse über das Original ins Modell; verbesserte Technik führt zu neuen Erkenntnissen über das Original 	<p><i>Wenn neue Informationen über die Kleinlibelle da sind, muss man sie halt neu bauen.</i></p>
	Ändern bei Veränderung des Originals <ul style="list-style-type: none"> Berücksichtigung von Veränderungen (z. B. Individualentwicklung) bzw. Weiterentwicklungen (z. B. evolutive Anpassung) des Originals als neue Informationen im Modell 	<p><i>Die Kleinlibelle könnte sich weiter entwickeln oder sie wird kleiner oder noch größer... [Das Modell muss verändert werden,] wenn sich vielleicht der Körper der Libelle an ein neues Klima anpassen muss. Tiere verändern sich ja auch von Zeit zu Zeit.</i></p>
III	Ändern bei Erkenntnissen aus Modellexperiment <ul style="list-style-type: none"> Anpassung des Modells an Erkenntnisse über das Original auf der Grundlage eines Modellexperiments/Falsifizieren der Hypothese, die dem Modell zugrunde liegt 	<p><i>Wenn man beim Test in einem Simulator merkt, dass sich das Modell ganz anders als gedacht verhält bzw. als die echte fliegt, dann könnte man etwas an den Gleitflächen verändern. Die Wissenschaftler hatten vielleicht eine andere Vermutung gehabt, woran das liegt.</i></p>

Anhang 2: Testmanual der Teilstudien 1 und 2

Manual für die Studie „Modellkompetenz im Biologieunterricht“

1. 15 Minuten vor dem eigentlichen Treffzeitpunkt vor Ort sein.

2. Einleitung

→ Name an die Tafel schreiben!

Mein Name ist _____ . Ich bin StudentIn an der Freien Universität Berlin und möchte im Rahmen meiner Doktorarbeit/Masterarbeit/Bachelorarbeit untersuchen, was ihr über biologische Modelle denkt.

Ich bin auf eure Mitarbeit angewiesen und hoffe, dass ihr mich tatkräftig unterstützt! Eure Antworten sind sehr wichtig für mich, weil sie Grundlage meiner Arbeit sind. Außerdem helfen diese Ergebnisse in der Zukunft, den Biologieunterricht noch besser zu machen. Deshalb bitte ich euch, die Fragen so gründlich wie möglich zu beantworten.

Ihr bekommt gleich einen Fragebogen ausgeteilt [Fragebogen hochhalten!]. Dieser besteht zunächst aus einem allgemeinen Teil. Bei diesem Teil sollt ihr ein paar allgemeine Angaben über euch machen! [Erste Seite des Fragebogens hochhalten!]

In dem Fragebogen befinden sich zudem Aufgaben zu verschiedenen biologischen Modellen [Seite im Fragebogen hochhalten!]. In dem Rahmen befindet sich eine Frage, die ihr bitte in ganzen Sätzen beantwortet. Bitte formuliert ganze Sätze, damit wir eure Aussagen später lesen und nachvollziehen können. Schreibt bitte so viel, wie euch einfällt. Je mehr desto besser!

Es kann sein, dass Themen angesprochen werden, die ihr noch nicht im Unterricht hatten. Beantwortet bitte trotzdem die Frage. Wenn ihr etwas an der Aufgabe nicht versteht, dann notiert dies bitte in der Aufgabe und meldet euch.

Eure Meinung und Ideen sind mir sehr wichtig. Richtige oder falsche Aussagen gibt es nicht. Bei der Bearbeitung könnt ihr also keine Fehler machen.

Die Fragebögen sind natürlich anonym. Das heißt, niemand erfährt, was ihr geantwortet habt und ihr bekommt auch keine Note dafür. Auch eure Lehrer erfahren nicht, wie ihr geantwortet habt.

Die Beantwortung des Fragebogens dauert ca. 40 Minuten.

Abschauen lohnt sich nicht, da ihr verschiedene Fragebögen bekommt und mich eure persönlichen Ideen und Meinungen interessieren!

Gibt es noch Fragen?

Viel Spaß beim Bearbeiten des Fragebogens!

3. Allgemeine Hinweise

- Bitte keine Hilfestellungen beim Lösen der Aufgaben geben!
- Wenn Fragen oder Probleme auftreten, bitte auf dem Protokoll die Frage und eure Antwort notieren. Bitte notiert auch, ob eure Antwort hilfreich war.
- Schülerinnen und Schüler sollen bei Verständnisschwierigkeiten die Aufgaben kommentieren.

4. Während der Testphase

- a. Protokoll ausfüllen (siehe Anhang)
- b. Auf Fragen der Schülerinnen und Schüler eingehen, aber keine Lösungen oder Tipps geben!

5. Nach der Testphase

- a. Bei Schülerinnen und Schülern und Lehrkräften für ihre Unterstützung bedanken!

Anhang 3: Konstruktionsanleitungen für Aufgaben im offenen Antwortformat

Anleitung zur Aufgabenkonstruktion für die Teilkompetenz *Eigenschaften von Modellen*

Eigenschaften von Modellen (Niveau I bis III)

Theoriebezug	<ul style="list-style-type: none"> - Schilderung der Beziehung zwischen Modell und Original - Perspektiven auf Modelle als Medien und als Methode: Modelle als Kopien (Niveau I), idealisierte Repräsentationen (Niveau II) oder theoretische Rekonstruktionen (Niveau III)
Struktur des Aufgabestamms	<ol style="list-style-type: none"> 1. Hinweis geben, dass das Original nicht direkt beobachtbar und nur mit Hilfsmitteln erkennbar ist (z. B. Röntgengerät, Mikroskop). 2. Daten beschreiben, die die Grundlage für die Konstruktion des Modells sind. 3. Grafische Darstellung des Prozesses der Entstehung des Modells: Bilder vom Original bzw. von den Untersuchungsergebnissen klar als Datengrundlage kennzeichnen. 4. Abbildung vom Modell zeigen.
Standardisierter Stimulus	<ul style="list-style-type: none"> - Beschreibe, inwieweit das Modell des [Originals] so aussieht wie das [Original]. Begründe deine Antwort!
Konstruktionshinweise	<ul style="list-style-type: none"> - Die gezeigten Originale stellen nicht selbst Modelle dar: Beispielsweise anstatt eines zusammengesetzten Knochenfundes gefundene Knochenreste oder Fossilien zeigen. - Nicht zu komplexe oder abstrakte Abbildungen des Originals wählen. Röntgenbilder beispielsweise sind weniger gut geeignet. - Wenn möglich nur ein Modell und passend dazu die jeweiligen Abbildungen des Originals zeigen, damit die Lernenden Modell und Original vergleichen. Werden mehrere Modelle gezeigt, neigen Lernende dazu, die Modelle miteinander statt das Original mit einem Modell zu vergleichen. - Original und Modell unterscheiden sich sichtbar voneinander bzw. grenzen sich ab, damit ein Interpretationsspielraum – auch für Niveau III – möglich ist.
Geeignete Kontexte	<ul style="list-style-type: none"> - Modelle, deren Originale nicht direkt beobachtbar sind, zum Beispiel Mikrokosmos, Evolution

Anleitung zur Aufgabenkonstruktion für die Teilkompetenz *Alternative Modelle*

Alternative Modelle (Niveau I bis III)

Theoriebezug	<ul style="list-style-type: none">- Begründungen für das Vorhandensein verschiedener Modelle zu einem Original- Perspektiven auf Modelle als Medien und als Methode: Gründe beziehen sich auf das Modellobjekt (Niveau I), auf den gelegten Fokus auf das Original (Niveau II) oder auf Hypothesen (Niveau III)
Struktur des Aufgabestamms	<ol style="list-style-type: none">1. Hinweis geben, dass das Original nicht direkt beobachtbar und nur mit Hilfsmitteln erkennbar ist (z. B. Röntgengerät, Mikroskop).2. Daten beschreiben, die die Grundlage für die Konstruktion der verschiedenen Modelle sind.3. Einen deutlichen Hinweis geben (z. B. fett markiert), dass diese Modelle zu einem Original sind.4. Grafische Darstellung des Prozesses der Entstehung der Modelle: Bilder vom Original bzw. von den Untersuchungsergebnissen präsentieren und diese als Datengrundlage klar kennzeichnen.5. Abbildungen der drei Modelle zeigen und nochmals darauf hinweisen, dass diese zu einem Original sind.
Standardisierter Stimulus	<ul style="list-style-type: none">- Erkläre, warum es zu diesem einen [Original] drei verschiedene Modelle gibt. Begründe deine Antwort!
Konstruktionshinweise	<ul style="list-style-type: none">- Im Stamm an mehreren Stellen deutlich kennzeichnen, dass die Modelle von einem Original sind.- Minimum drei Modelle zeigen, die verschiedene Materialeigenschaften (z. B. Farbe, Dimensionen), verschiedene inhaltlichen Schwerpunkte und verschiedene Hypothesen zeigen.- Wenn möglich keine Realobjekte verwenden, weil Lernende sonst nicht plausibel auf Niveau I (Unterschiede zwischen den Modellobjekten) antworten können.
Geeignete Kontexte	<ul style="list-style-type: none">- Modelle, deren Originale nicht direkt beobachtbar sind, zum Beispiel Mikrokosmos.

Anleitung zur Aufgabenkonstruktion für die Teilkompetenz *Zweck von Modellen*

Zweck von Modellen (Niveau I bis III)

Theoriebezug	<ul style="list-style-type: none"> - Schilderung der Beziehung zwischen Modell und Original - Perspektiven auf Modelle als Medien und als Methode: Beschreibung von Merkmalen des Originals (Niveau I), Erklärung von Zusammenhängen im Original (Niveau II) oder Voraussage von neuen Erkenntnissen/Beschreibungen von Hypothesen (Niveau III)
Struktur des Aufgabenstamms	<ol style="list-style-type: none"> 1. Verweis auf das Foto zum Modell geben. 2. Abbildung des Modells zeigen.
Standardisierter Stimulus	<ul style="list-style-type: none"> - Modelle werden für einen bestimmten Zweck entwickelt. Beschreibe, welchen Zweck dieses [Original]-Modell haben kann. Begründe deine Antwort!
Konstruktionshinweise	<ul style="list-style-type: none"> - Wenn möglich keine weiteren Informationen zum Original oder Modell im Aufgabenstamm erwähnen, weil diese immer eine Zweckbeschreibung beinhalten und somit die Lernenden beeinflusst. - Wenn möglich nicht mehrere Abbildungen zu einem Modell zeigen, weil dies Lernende häufig verwirrt und zudem Zusammenhänge suggeriert werden, die wiederum Antworten auf Niveau I (Beschreibung von Sachverhalten) verhindern. - Modelle wählen, die ohne weitere Informationen im Stamm verständlich sind. Kontexte wie Nahrungsbeziehungen oder enzymatische Prozesse bedürfen wegen der verwendeten Symbole einer näheren Erklärung und sind daher ohne weitere Informationen von den Lernenden weniger gut zu verstehen. - Kontexte (z. B. Nahrungsbeziehungen, Mundapparat) vermeiden, die bereits einen deutlichen Zusammenhang darstellen. Lernende können sonst keine plausiblen Aussagen auf Niveau I (Beschreibung von Sachverhalten) formulieren.
Geeignete Kontexte	<ul style="list-style-type: none"> - Realobjekte (z. B. Ökosysteme), die im Ganzen nicht direkt beobachtbar sind. - Systemische Modelle (z. B. Ökosysteme), die wenig vereinfacht sind.

Anleitung zur Aufgabenkonstruktion für die Teilkompetenz *Testen von Modellen*

Testen von Modellen

Theorie- bezug	<ul style="list-style-type: none"> - Vorgehensweisen bei der Testung von Modellen - Perspektiven auf Modelle als Medien und als Methode: Überprüfung des Modellobjekts (Niveau I), Vergleich von Modell und Original (Niveau II) oder Überprüfung von Hypothesen über das Original mit dem Modell (Niveau III)
Struktur des Aufgaben- stamms	<ol style="list-style-type: none"> 1. Die Eigenschaft des Originals nennen, die untersucht werden soll. 2. Beschreiben, dass diese Eigenschaft mithilfe eines Modells näher untersucht wird (Zweck des Modells). 3. Abbildung vom Original und dem passenden Modell zeigen.
Standardi- sierter Sti- mulus	<ul style="list-style-type: none"> - Erkläre ausführlich, wie man überprüfen kann, ob das Modell des [Originals] seinen Zweck erfüllt.
Konstruktio- nshinweise	<ul style="list-style-type: none"> - Das „wie“ im Stimulus deutlich kennzeichnen, damit nicht nur der Zweck des Modells aus dem Aufgabenstamm wiederholt wird. - Den konkreten Zweck des Modells im Aufgabenstamm nennen und im Stimulus abstrakt darauf hinweisen („Zweck erfüllt“), um den Lernenden nicht zu beeinflussen. - Abbildung vom Original zeigen, um der Beschreibung von Niveau II gerecht zu werden. - Wenn möglich das Zeigen mehrere Modelle vermeiden, weil dies Lernende verwirrt und das Original weniger beachtet wird. - Wenn möglich keine Realobjekte verwenden, weil diese keinen plausiblen Vergleich mit dem Original ermöglichen. - Wenn möglich keine Schwarz-Weiß-Zeichnungen verwenden, weil Lernende keine plausiblen Antworten auf Niveau I und Niveau III generieren können.
Geeignete Kontexte	<ul style="list-style-type: none"> - Technische Modelle, die plausibel eine Testung des Modellobjekts, ein Parallelisieren mit dem Original und eine hypothesengeleitete Testung erlauben.

Anleitung zur Aufgabenkonstruktion für die Teilkompetenz *Ändern von Modellen*

 Ändern von Modellen

Theorie- bezug	<ul style="list-style-type: none"> - Begründungen für Veränderungen am Modell - Perspektiven auf Modelle als Medien und als Methode: Veränderungen am Modellobjekt (Niveau I), Veränderungen aufgrund neuer Erkenntnisse über das Original (Niveau II) oder Veränderungen aufgrund der Falsifikation der Hypothese über das Original (Niveau III)
Struktur des Aufgaben- stamms	<ol style="list-style-type: none"> 1. Die Eigenschaft des Originals nennen, die untersucht werden soll. 2. Beschreiben, dass ein Modell des genannten Originals gebaut wurde. 3. Abbildung vom Modell zeigen
Standardi- sierter Sti- mus	<ul style="list-style-type: none"> - Nenne Gründe, warum dieses Modell eines [Originals] verändert werden könnte!
Konstruktio- nshinweise	<ul style="list-style-type: none"> - Eine Abbildung des Originals muss mit Blick auf die Theorie nicht notwendigerweise gezeigt werden. - Wenn möglich keine Veränderungsprozesse des Modells zeigen, weil sonst basale Perspektiven (keine Änderung des Modells) verhindert und nicht erfasst werden können. - Wenn möglich keine einfachen Modelle zum Beispiel Zanonias heranziehen, weil diese die Lernenden nicht dazu anregen, vielfältige Ideen für Veränderungen zu generieren.
Geeignete Kontexte	<ul style="list-style-type: none"> - Technische Modelle, die eine plausible Testung und damit Veränderung des Modells erlauben. - Systemische Modelle (z. B. Mundapparat), die wenig vereinfacht sind.

Anhang 4: Testmanual der Hauptstudie

Manual für die Studie „Modellkompetenz im Biologieunterricht“

1. 15 Minuten vor dem eigentlichen Treffzeitpunkt vor Ort sein.

2. Einleitung

Name an die Tafel schreiben!

Unsere Namen sind _____. Wir sind Doktoranden an der Freien Universität Berlin und möchte im Rahmen unserer Doktorarbeit untersuchen, was ihr über biologische Modelle denkt.

Wir sind auf eure Mitarbeit angewiesen und hoffen, dass ihr uns tatkräftig unterstützt!

Eure Antworten sind sehr wichtig für uns, weil sie Grundlage unserer Arbeit sind. Außerdem helfen diese Ergebnisse in der Zukunft, den Biologieunterricht noch besser zu machen. Deshalb bitten wir euch, alle Fragen zu beantworten und diese so gründlich wie möglich zu beantworten.

Ihr bekommt verschiedene Fragebögen ausgeteilt. Manchmal müsst ihr ganze Sätze formulieren und manchmal könnt ihr die Fragen durch Ankreuzen beantworten. Wichtig ist es, dass ihr alle Aufgaben beantwortet.

In den verschiedenen Aufgaben geht es um eure Meinung zu verschiedenen biologischen Modellen. Richtige oder falsche Aussagen gibt es nicht. Bei der Bearbeitung könnt ihr also keine Fehler machen. Es kann sein, dass Themen angesprochen werden, die ihr noch nicht im Unterricht hattet. Trotzdem interessiert uns eure persönliche Meinung dazu! Auch wenn ihr euch mal nicht sicher seid, bitte trotzdem die Aufgabe beantworten. Wichtig ist zudem, dass ihr die Fragen ernsthaft beantwortet, denn nur so haben wir die Möglichkeit, gute Ergebnisse zu bekommen. Abschauen lohnt sich nicht, da ihr verschiedene Fragebögen bekommt und uns eure persönlichen Ideen und Meinungen interessieren!

Nun zur Vorgehensweise: Die Befragung dauert insgesamt 90 Minuten. Dazu braucht ihr nur einen Stift. Bitte alle anderen Dinge vom Tisch räumen. Jeder von euch bekommt einen Stapel mit verschiedenen Heften. Bitte immer mit Teil 1 anfangen, wenn ihr fertig seid, diesen zur Seite legen und den nächsten Teil bearbeiten. Bitte nur in dieser Reihenfolge die Hefte bearbeiten, wie sie vor euch liegen.

Wenn ihr alle Hefte bearbeitet habt, werden wir alle gemeinsam noch einen Lesetest und kurzen Intelligenztest mit euch machen.

Gibt es noch Fragen zu dieser Vorgehensweise?

Wir danken euch schon im Voraus für eure Mitarbeit!

3. Allgemeine Hinweise zur Strukturierung

1. Schritt: Bearbeitung der Testpakete

→ **4 Teile für jeden Schüler austeilen:** Darauf achten, dass jedes Heft dieselbe Schülernummer hat; dabei teilt eine Person aus, eine andere Person passt auf, dass die Testhefte nicht ausgetauscht werden oder in andere Testhefteile geschaut wird.

Sprechtext:

Jeder von euch bekommt vier Hefte, die ihr bitte nacheinander bearbeitet. Ihr beginnt mit Teil 1, wenn ihr fertig seid, diesen zur Seite legen und den nächsten Teil bearbeiten. Bitte nur in der Reihenfolge die Hefte bearbeiten, wie sie vor euch liegen.

→ **Hinweise:** Wenn ein Schüler früher fertig ist, ein Rätsel geben, damit er die anderen Lernenden nicht stört. Während der Bearbeitung sollten die Fragen von Schülern beantwortet werden und ein Sitzplan mit Sitznummern angefertigt werden.

2. Schritt: LGVT

- Folie mit Anweisung auflegen und den weiteren Ablauf erklären.
- LGVT an alle austeilen.
- Start des Lesetests ansagen und Zeit stoppen (4 Minuten, Stoppuhr mitbringen).
- Danach Lesetests wieder einsammeln.

3. Schritt: KFT

- Folie mit Anweisung auflegen und den weiteren Ablauf erklären.
- KFT an alle austeilen
- Start des KFTs ansagen und Zeit stoppen (8 Minuten, Stoppuhr mitbringen).
- Danach KFTs wieder einsammeln
- Überprüfen, ob Testform A oder B angekreuzt wurde!

4. Während der Testphase

- Auf Fragen der Schülerinnen und Schüler eingehen, aber keine Lösungen oder Tipps geben!
- Auffälligkeiten auf dem Protokoll vermerken!

5. Nach der Testphase

- Bei Schülerinnen und Schülern und Lehrkräften für ihre Unterstützung bedanken!

Anhang 5: Kennwerte der zweidimensionalen IRT- Skalierung

Kennwerte der zweidimensionalen Raschskalierung auf Aufgabenebene

Aufgabe*	Bearbeitungen	Trennschärfe ^a	Schwierigkeit ^b	wMNSQ	t-Werte
EM1: Neandertaler	230	.56	0.468	0.99	0.0
EM2: Dinosaurier	232	.46	0.308	1.05	0.5
EM3: Biomembran	235	.43	0.914	1.00	0.0
A1: Biomembran	235	.85	0.527	0.98	-0.2
A2: Speiseröhre	235	.74	0.879	0.97	-0.4
A3: Zunge	235	.75	0.485	0.98	-0.3
Z1: Wald	235	.69	-1.234	0.98	-0.2
Z2: Meer	235	.79	-0.935	1.02	0.2
Z3: Darm	236	.63	-0.593	0.96	-0.5
T1: Libelle	237	.68	-0.222	0.99	-0.1
T2: Zanonía	237	.65	-0.088	1.01	0.1
T3: Käfer	237	.66	0.038	0.99	-0.1
Ä1: Libelle	237	.67	1.148	0.98	-0.2
Ä2: Krabbe	238	.88	1.107	1.05	0.5
Ä3: Mund	237	.75	0.984	1.00	0.0

Anmerkungen. * Die Buchstaben kennzeichnen die Teilkompetenz der Aufgabe (EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen). Die Zahl kennzeichnet die Nummer der Aufgabe, zum Beispiel Z1 = Aufgabe 1, der Teilkompetenz **Zweck von Modellen**.

^a Für die Berechnung der Trennschärfen wurden die Aufgaben pro Teilkompetenz eindimensional skaliert.

^b Die Werte der Aufgabenschwierigkeit beruhen auf Mittelwerten der Thresholdparameter.

Mittelwerte der Personenfähigkeiten auf Ebene der Antwortkategorien

Aufgabe*	Mittelwerte bei den Personenfähigkeiten auf Ebene der Antwortkategorien (0, 1, 2 und 3) ^a			
	0	1	2	3
EM1	-0.93	-0.20	0.74	1.13
EM2	-1.13	-0.20	0.79	0.65
EM3	-0.70	0.05	0.83	1.66
A1	-0.73	-0.15	0.71	1.28
A2	-0.71	0.01	0.67	1.44
A3	-0.83	-0.38	0.60	1.06
Z1	-1.59	-0.44	-0.11	0.58
Z2	-1.15	-0.61	0.12	0.50
Z3	-1.31	-0.52	-0.03	0.65
T1	-1.20	-0.39	0.21	0.67
T2	-0.53	-0.38	0.14	0.66
T3	-1.11	-0.43	0.23	0.16
Ä1	-0.86	-0.29	0.21	1.45
Ä2	-0.68	-0.38	0.22	1.36
Ä3	-0.57	-0.52	0.29	0.09

Anmerkungen. * Die Buchstaben kennzeichnen die Teilkompetenz der Aufgabe (EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen). Die Zahl kennzeichnet die Nummer der Aufgabe, zum Beispiel Z1 = Aufgabe 1, der Teilkompetenz **Zweck von Modellen**.

^a Unter Antwortkategorie 0 sind unbeantwortete Aufgaben, nicht mit der Frage korrespondierende Antworten sowie Antworten auf dem basalen Niveau subsumiert. Die Antwortkategorien 1, 2 und 3 entsprechen den Niveaus des Kompetenzmodells

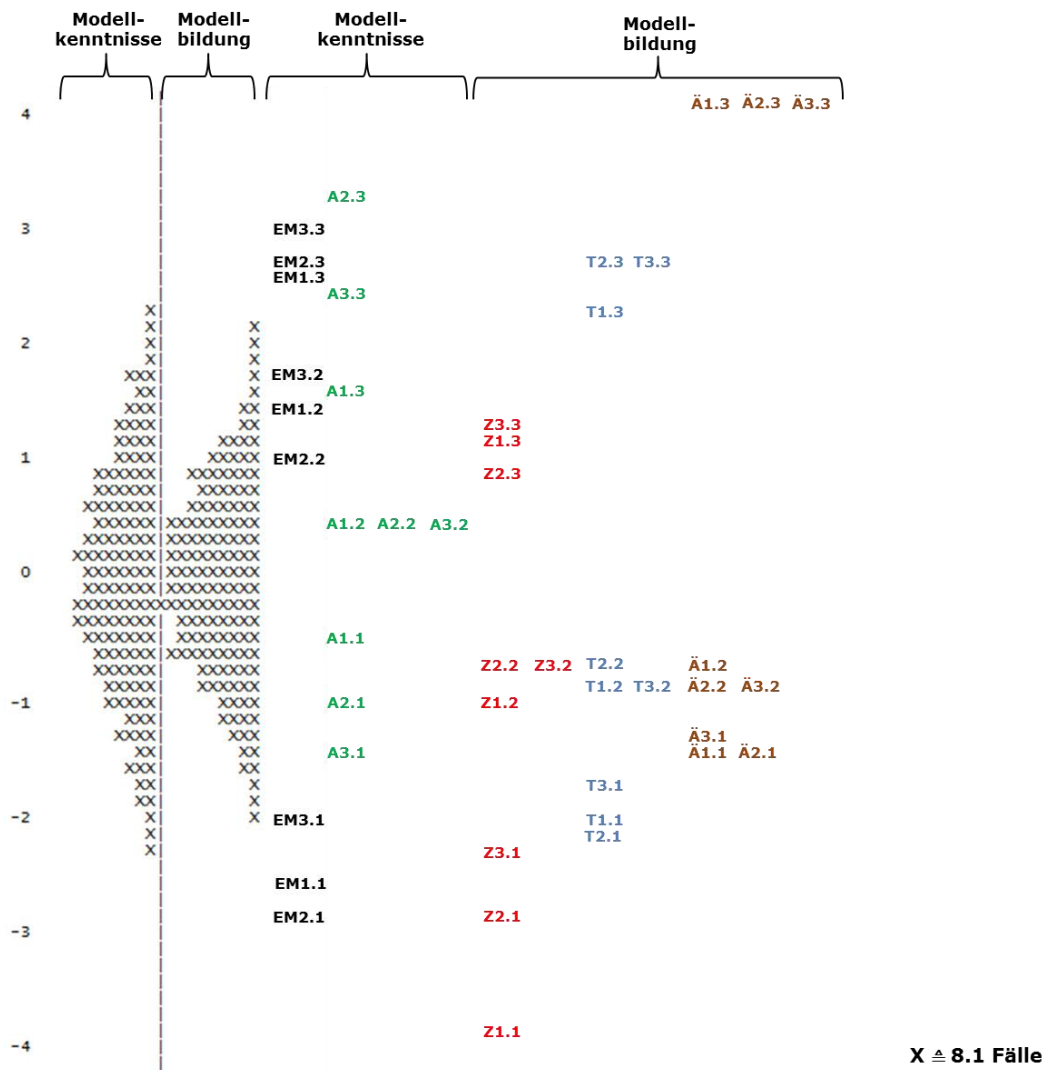
Kennwerte der zweidimensionalen Raschskalierung auf Ebene der Antwortkategorien

Auf- gabe*	Häufigkeiten (%)				wMNSQ				t-Werte				Punktbiseriale Korrelationen				Item-Deltas (δ)			Item Thresholds (τ) ^a		
	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3	δ_1	δ_2	δ_3	τ_1	τ_2	τ_3
EM1	9.13	71.30	14.78	4.78	1.03	0.97	0.95	1.04	0.2	-0.4	-0.4	0.2	-0.12	-0.05	0.10	0.11	-2.66	1.68	2.20	-2.67	1.48	2.59
EM2	6.90	65.52	22.84	4.74	0.95	1.02	1.02	1.06	-0.2	0.4	0.3	0.3	-0.09	-0.08	0.15	0.00	-2.89	1.25	2.56	-2.90	1.07	2.76
EM3	14.89	69.36	12.77	2.98	0.98	0.98	0.98	1.02	-0.1	-0.3	-0.1	0.1	-0.17	0.01	0.15	0.05	-1.99	2.09	2.64	-2.01	1.77	2.99
A1	35.74	28.09	22.13	14.04	0.97	1.00	0.97	1.03	-0.4	0.0	-0.4	0.3	-0.28	-0.01	0.22	0.14	-0.17	0.47	1.29	-0.53	0.50	1.60
A2	24.26	36.60	35.32	3.83	0.98	0.99	0.95	1.04	-0.2	-0.4	-0.9	0.2	-0.20	-0.07	0.27	-0.04	-0.89	0.30	3.23	-1.11	0.47	3.28
A3	19.15	41.70	31.91	7.23	0.97	0.99	0.96	1.08	-0.3	-0.4	-0.8	0.5	-0.10	-0.18	0.23	0.08	-1.34	0.44	2.36	-1.48	0.45	2.48
Z1	1.28	25.11	49.79	23.83	1.04	1.00	0.98	0.97	0.2	-0.1	-0.8	-0.3	-0.13	-0.20	0.12	0.10	-3.79	-0.94	1.02	-3.84	-1.00	1.14
Z2	3.40	27.66	41.70	27.23	1.03	0.98	0.98	1.04	0.2	-0.2	-0.6	0.5	-0.07	-0.23	0.20	0.03	-2.88	-0.65	0.73	-2.98	-0.74	0.92
Z3	5.93	26.27	47.46	20.34	0.97	0.98	0.98	1.00	-0.1	-0.3	-0.8	0.1	-0.14	-0.26	0.13	0.20	-2.21	-0.78	1.21	-2.39	-0.71	1.32
T1	7.59	19.41	62.87	10.13	0.96	1.00	1.00	1.02	-0.2	0.0	0.0	0.2	-0.20	-0.04	0.06	0.13	-1.65	-1.31	2.30	-2.05	-0.95	2.32
T2	7.17	24.89	61.18	6.75	1.03	0.97	0.99	1.02	0.2	-0.4	-0.2	0.2	-0.09	-0.22	0.18	0.12	-1.93	-1.02	2.69	-2.20	-0.78	2.71
T3	10.13	16.88	66.24	6.75	0.99	1.01	0.96	1.01	-0.1	0.1	-0.7	0.1	-0.23	-0.10	0.25	-0.05	-1.16	-1.47	2.75	-1.74	-0.92	2.76
Ä1	14.77	18.57	66.24	0.42	0.95	1.00	1.00	1.04	-0.4	0.1	0.0	0.4	-0.20	-0.01	0.17	-0.04	-0.85	-1.33	5.62	-1.47	-0.70	5.62
Ä2	13.87	17.23	68.49	0.42	1.07	0.99	1.02	1.01	0.5	-0.1	0.2	0.3	-0.08	-0.15	0.17	0.10	-0.85	-1.46	5.64	-1.52	-0.80	5.64
Ä3	15.19	14.35	69.62	0.84	1.01	1.00	0.98	1.02	0.1	0.0	-0.2	0.3	-0.18	-0.17	0.27	-0.01	-0.49	-1.58	5.02	-1.32	-0.75	5.02

Anmerkungen. Unter Antwortkategorie 0 sind unbeantwortete Aufgaben, nicht mit der Frage korrespondierende Antworten sowie Antworten auf dem basalen Niveau subsumiert. Die Antwortkategorien 1, 2 und 3 entsprechen den Niveaus des Kompetenzmodells.

* Die Buchstaben kennzeichnen die Teilkompetenz der Aufgabe (EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen). Die Zahl kennzeichnet die Nummer der Aufgabe, zum Beispiel Z1 = Aufgabe 1, der Teilkompetenz **Zweck von Modellen**.

^a Diejenige Stelle auf der Skala, an der die Wahrscheinlichkeit die Antwortkategorie k oder einer höhere zu erzielen 50 Prozent beträgt.



Wright Map der zweidimensionalen Skalierung mit den Personenfähigkeiten (links) und den Thurstonian Thresholds (rechts)¹

Varianzen und EAP/PV-Reliabilitäten der zweidimensionalen Raschskalierung

Kennwerte	Modellkenntnisse	Modellbildung
Varianz	1.136	0.750
EAP/PV-Reliabilität	.341	.360

¹ Die Buchstaben kennzeichnen die Teilkompetenz der Aufgabe (EM = Eigenschaften von Modellen, A = Alternative Modelle, Z = Zweck von Modellen, T = Testen von Modellen, Ä = Ändern von Modellen). Die erste Zahl kennzeichnet die Nummer der Aufgabe und die zweite Zahl den Thurstonian Threshold, z. B. Z1.1 = Aufgabe 1, Threshold 1 (Schwelle an der es wahrscheinlicher ist, Kategorie 1 anstatt Kategorie 0 zu wählen) der Teilkompetenz **Zweck von Modellen**.

Anhang 6: Publikationsliste

Beiträge zum Forschungsthema (mit Review-Verfahren)

Grünkorn, J., Upmeier zu Belzen, A. & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education*, 1-34. doi: 10.1080/09500693.2013.873155

Grünkorn, J. & Krüger, D. (2012). Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells zur Modellkompetenz. In U. Harms & F. X. Bogner (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (Band 5, S. 9-27). Innsbruck: Studienverlag.

Grünkorn, J., Upmeier zu Belzen, A. & Krüger, D. (2011). Design and test of open-ended tasks to evaluate a theoretical structure of model competence. In A. Yarden & G.S. Carvalho (Eds.), *Authenticity in Biology Education. Benefits and Challenges* (S. 53-65). Braga, Portugal: CIEC Universidade do Minho.

Beiträge zum Forschungsthema

Grünkorn, J., Lotz, A. & Terzer, E. (2014). Erfassung von Modellkompetenz im Biologieunterricht. *Der mathematische und naturwissenschaftliche Unterricht*, 67 (3), 132-138.

Grünkorn, J., Hänsch, J., Upmeier zu Belzen, A., & Krüger, D. (2012). Determination of students' model competence using open-ended and hands-on tasks. In C. Bruguière, A. Tiberghien & P. Clément (Eds.), *E-Book Proceedings of the ESERA 2011 Conference: Science learning and Citizenship. Part 5* (co-ed Maurines, L., & Redfors, A.), (S. 39-45) Lyon, France: European Science Education Research Association.

Beiträge zu Unterrichtsthemen

Grünkorn, J. & Hanauer, N. (2013a). Fischhaut [Materialien]. *Grundschule*, 45(6), Beilage, XIII-XVI.

Grünkorn, J. & Hanauer, N. (2013b). Schwimmen wie Fische im Wasser – Untersuchung der Schleimschicht von Fischen. *Grundschule*, 45 (6), 28-29.

Grünkorn, J. & Hanauer, N. (2013). Einer für alle - alle für einen. In P. Schmiemann & J. Mayer (Hrsg.), *Experimentieren Sie! Biologieunterricht mit Aha-Effekt. Selbständiges, kompetenzorientiertes Erarbeiten von Lehrplaninhalten* (S. 78-80). Berlin: Cornelsen Scriptor.

Grünkorn, J. & Fleige, J. (2012). Bau und Funktion der Fischhaut. In J. Fleige, A. Seegers, A. Upmeyer zu Belzen & D. Krüger (Hrsg.), *Modellkompetenz im Biologieunterricht Klasse 7-10. Phänomene begreifbar machen – in 11 komplett ausgearbeiteten Unterrichtseinheiten* (S. 23-28). Donauwörth: Auer.

Beiträge in Tagungsbänden (mit Review-Verfahren)

Grünkorn, J. & Krüger, D. (2011). Empirische Überprüfung eines Kompetenzmodells zur Modellkompetenz mit offenen Aufgaben [Abstract]. In F. X. Bogner (Hrsg.), *Internationale Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO*. (S. 14-15). Bayreuth: Universität Bayreuth.

Grünkorn, J., Hänsch, J., Upmeyer zu Belzen, A. & Krüger, D. (2011). **Determination of students' model competence using open-ended and hands-on tasks** [Abstract]. 9. Tagung der European Science Education Research Association (Esera). Lyon. In C. Bruguière (Hrsg.), *Ninth Conference of the European Science Education Research Association (ESERA)*. (S. 139). Lyon: Lyon Ingenierie Projets.

Grünkorn, J. & D. Krüger (2011). Model competence in biology education – evaluation of a theoretical structure using open-ended tasks [Abstract]. **14. Tagung der European Association for Research on Learning and Instruction (Earli)**, 494-495. Zugriff am 01.10.2012. Verfügbar unter http://www.earli2011.org/media/Documents_EARLI2011/BookofAbstractsandSummaries.pdf

Grünkorn, J. & Krüger, D. (2011). Überprüfung der Struktur eines Kompetenzmodells zur Modellkompetenz mit offenen Aufgaben [Abstract]. In C. Florian, P. Schmiemann & A. Sandmann (Hrsg.). **13. Frühjahrsschule der Fachsektion Biologiedidaktik im VBIO**. (S. 69-70). Mülheim a. d. R.: Universität Duisburg-Essen.

Krüger, D., Upmeyer zu Belzen, A., Grünkorn, J., Terzer, E., Krell, M. & Hänsch, J. (2011). Mit Modellen zum Wissenschaftsverständnis – Strukturie-

rung, Erfassung und Förderung von Modellkompetenz [Abstract]. Schwerpunkttagung der Gesellschaft für die Didaktik der Chemie und Physik (GDGP) gemeinsam mit der Fachsektion Didaktik der Biologie des VBio - Lehren und Lernen über die Natur der Naturwissenschaften , 2-3. Zugriff am 1.03.2011. Verfügbar unter http://aecc.univie.ac.at/fileadmin/user_upload/kompetenzzentrum_aeccp/GDCP_Schwerpunkt/Programm___Abstracts_110127.pdf

Grünkorn, J. & Krüger, D. (2010). Model competence in biology education – validation of a theoretical model of model competence using open-ended items [Abstract]. In A. Yarden, D. J. Boerwinkel, M. Ekborg, M. Reiss, V. Zogza, G. Carvalho, et al. (Hrsg.), *Eight Conference of European Researchers in Didactics of Biology (ERIDOB)*. (S. 115). Braga: Universidade do Minho.

Grünkorn, J., Upmeyer zu Belzen, A. & Krüger, D. (2009). Diagnose von Modellkompetenz im Biologieunterricht – Validierung eines Kompetenzmodells mit offenen und halboffenen Aufgaben [Abstract]. In U. Harms, F.X. Bogner, D. Graf, H. Gropengießer, D. Krüger, J. Mayer, et al. (Hrsg.), *Internationale Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO*. (S. 172-175). Kiel: IPN.

Grünkorn, J., Upmeyer zu Belzen, A. & Krüger, D. (2009). Modellkompetenz im Biologieunterricht - Überprüfung eines Kompetenzmodells mit offenen und geschlossenen Aufgaben [Abstract]. In S. Hof, K. Kremer, A. Upmeyer zu Belzen & D. Krüger (Hrsg.), *11. Frühjahrsschule der Fachsektion Biologiedidaktik im VBIO*. (S. 18-19). Gießen: Justus-Liebig-Universität Gießen.

Beitrag im Tagungsband

Grünkorn, J. (2010). **School's Next Top MODELL [Abstract]**. 8. Herbst-Kongresses des Berlin-Brandenburger Landesvereins MNU, 16. Zugriff am 01.12.2010. Verfügbar unter http://www.mnu-berlin.de/kongress10/pdf/Tagungsheft_MNUBB_10.pdf

Vorträge (* eingeladener Vortrag)

- *Grünkorn, J. (2013, Juni). Kompetenzen und die Herausforderungen, sie zu messen. Vortrag auf der Veranstaltung des Verbundprojekts „**Lernen vor Ort**“ der Fachstelle Bildung des Rheingau-Taunus-Kreises und des Büros „**Lernende Netzwerk Region**“ der vhs Rheingau-Taunus e.V. mit dem Thema „**Kompetenzmessung und -erfassung in pädagogischen Handlungsfeldern**“, Bad Schwalbach.
- Grünkorn, J. & Krüger, D. (2011, September). Empirische Überprüfung eines Kompetenzmodells zur Modellkompetenz mit offenen Aufgaben. Vortrag auf der Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, Bayreuth.
- Grünkorn, J., Hänsch, J., Upmeier zu Belzen, A. & Krüger, D. (2011, September). **Determination of students' model** competence using open-ended and hands-on Tasks. Vortrag auf der 9. Tagung der European Science Education Research Association (Esera), Lyon.
- Grünkorn, J. & Krüger, D. (2011, September). Model competence in biology education – evaluation of a theoretical structure using open-ended tasks. Vortrag auf der 14. Tagung der European Association for Research on Learning and Instruction (Earli), Exeter.
- Upmeier zu Belzen, A., Grünkorn, J., Terzer, E., Krell, M. & Hänsch, J. (2011, Februar). Mit Modellen zum Wissenschaftsverständnis – Strukturierung, Erfassung und Förderung von Modellkompetenz. Vortrag auf der Schwerpunkttagung der Gesellschaft für die Didaktik der Chemie und Physik (GDPC) gemeinsam mit der Fachsektion Didaktik der Biologie des VBio – Lehren und Lernen über die Natur der Naturwissenschaften, Wien.
- Grünkorn, J. & Krüger, D. (2011, April). Überprüfung der Struktur eines Kompetenzmodells zur Modellkompetenz mit offenen Aufgaben. Vortrag auf der 13. Frühjahrsschule der Fachsektion Biologiedidaktik im VBIO, Mülheim a. d. R..
- Grünkorn, J. (2010, September). **School's Next Top** MODELL. Vortrag auf dem 8. Herbst-Kongress des Berlin-Brandenburger Landesvereins MNU, Berlin.

Poster

Grünkorn, J. & Krüger, D. (2010, Juli): Development and evaluation of open-ended tasks to test model competence. Poster auf der Eighth Conference of European Researchers in Didactics of Biology (ERIDOB), Braga.

Grünkorn, J., Upmeyer zu Belzen, A. & Krüger, D. (2009, September). Diagnose von Modellkompetenz im Biologieunterricht – Ergebnisse einer Voruntersuchung zur Dimension Modellkenntnisse. Poster auf der Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, Kiel.

Grünkorn, J., Upmeyer zu Belzen, A. & Krüger, D. (2009, März). Modellkompetenz im Biologieunterricht - Validierung eines Kompetenzmodells mit offenen und halboffenen Aufgaben. Poster auf der 11. Frühjahrsschule der Fachsektion Biologiedidaktik im VBIO, Marburg.

Workshops (* eingeladener Workshop)

*Grünkorn, J. & Terzer, E. (2013, Dezember). *Empirische Beschreibung von Modellkompetenz im Biologieunterricht*. Workshop für Studentinnen und Studenten der Georg-August-Universität Göttingen im Master of Education, Göttingen.

*Terzer, E. & Grünkorn, J. (2012, November). *Aktuelle Forschung zu Modellkompetenz im Biologieunterricht*. Workshop für Studentinnen und Studenten der Georg-August-Universität Göttingen im Master of Education, Göttingen.

*Terzer, E. & Grünkorn, J. (2011, Dezember). *Modellkompetenz im Biologieunterricht – Perspektiven aus der Forschung*. Workshop auf dem Lehrgang der Landesakademie für Fortbildung und Personalentwicklung an Schulen mit dem Thema „Biologie: Entwicklung von standardbasiertem und kompetenzorientiertem Unterricht im Fach Biologie für die Kursstufe“, Bad Wildbad.

*Grünkorn, J. & Terzer, E. (2011, November). *Modellkompetenz im Biologieunterricht*. Workshop für Studentinnen und Studenten der Georg-August-Universität Göttingen im Master of Education, Göttingen.