

Diagnose von Modellkompetenz

Deduktive Konstruktion und Selektion von geschlossenen Items

Moritz Krell & Dirk Krüger

Moritz.Krell@fu-berlin.de – Dirk.Krueger@fu-berlin.de

Freie Universität Berlin, Didaktik der Biologie,
Schwendenerstraße 1, 14195 Berlin

Zusammenfassung

Das Ziel dieses Projektes ist die Entwicklung eines für den schulischen Einsatz geeigneten Diagnoseinstruments für Modellkompetenz. Vorliegender Artikel stellt die Konstruktion und Selektion geschlossener Items vor, die im weiteren Verlauf des Projekts die Grundlage des Diagnoseinstruments bilden sollen. Auf der Basis eines Kompetenzmodells für Modellkompetenz werden deduktiv Items formuliert und inhaltlich validiert. Eine Bewertung der Items durch SchülerInnen und Studierende bildet den Ansatzpunkt für die statistische Auswertung und Selektion der Items. Die ausgewählten Items werden im weiteren Verlauf des Projekts sowohl qualitativ als auch quantitativ weitergehend untersucht.

Abstract

The aim of this project is to develop an instrument for the diagnostic analysis of model competence in school. Present article focuses on the construction und selection of closed test-items. In the further process of this project these items will be the fundament of the diagnostic instrument. Based on a model of model competence items are formulated deductively and content validated. A rating of the items by pupils and students serves as a starting point for the statistic analysis and selection of the items. The selected items will be explored qualitatively and quantitatively in additional investigations.

1 Einleitung

Modelle sind wesentlicher Bestandteil naturwissenschaftlicher Erkenntnisgewinnung. Sie dienen sowohl als Medium zur Darstellung als auch als Methode zur Gewinnung naturwissenschaftlichen Wissens. Die Bedeutung von Modellen für die Naturwissenschaft wird von GILBERT (1991) hervorgehoben, der Naturwissenschaft als Prozess der Modellbildung definiert. Diese Bedeutung von Modellen für Naturwissenschaft findet ihren Wiederhall in öffentlichen Zielvorgaben für allgemeinbildende Schulen (KMK, 2005). UPMEIER ZU BELZEN & KRÜGER (2010) haben ein Kompetenzmodell der Modellkompetenz entwickelt, welches als inhaltliche Orientierung für einen Unterricht mit und über Modelle und als Grundlage für die Diagnose von Modellkompetenz dient. Grundlage des Kompetenzmodells ist der Kompetenzbegriff nach WEINERT (2001), der neben kognitiven Facetten auch soziale, motivationale und volitionale Aspekte mit einschließt:

„Modellkompetenz umfasst die Fähigkeiten, 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, 49).

Vorliegendes Projekt zielt darauf ab, auf der Grundlage des Kompetenzmodells Diagnoseaufgaben für Modellkompetenz zu erarbeiten. Hierfür werden deduktiv Items entwickelt und in unterschiedliche geschlossene Aufgaben implementiert. Die Entwicklung verschiedener Aufgaben geschieht, um deren Ergebnisse statistisch vergleichen zu können. Das Projekt fokussiert durch die Verwendung geschlossener Aufgaben auf die Messung bzw. Diagnose des individuellen Modellverständnisses, im Sinne eines allgemein-konzeptuellen Verständnisses über Modelle. Modellverständnis ist notwendige Voraussetzung für Modellkompetenz (TERZER & UPMEIER ZU BELZEN, 2007).

2 Theorie

2.1 Deduktive Methodologie

Der Philosoph Popper hat mit seiner Arbeit ein logisch schlüssiges Vorgehen für wissenschaftliche Erkenntnisgewinnung erarbeitet: Die deduktive Methode. Ausgangspunkt ist die Überzeugung, dass jede menschliche Beobachtung bereits theoriegeleitet ist und folglich eine Erklärung bzw. Interpretation eines

Phänomens auf rein induktivem Wege schlechterdings nicht möglich ist (POPPER, 1963/2009). Hieraus zieht Popper den Schluss, dass Wissenschaft primär in der „deduktiven Methodik der Nachprüfung“ (POPPER, 1934/2005, S.6) von Theorien bestehen sollte. Diese besteht in einem Dreischritt: Aus einer Theorie werden deduktiv Folgerungen (bzw. „Prognosen“; POPPER, 1934/2005, S.9) abgeleitet, diese werden empirisch (und logisch) geprüft, um schließlich beurteilt zu werden. Dieses Urteil kann nur im Falle der Falsifikation einer Theorie endgültigen Charakter haben. Bewähren sich die Folgerungen dagegen empirisch (und logisch), kann lediglich von einer Plausibilisierung, niemals von einer Verifizierung, gesprochen werden.

Zwar wurde der Ansatz Poppers teilweise stark kritisiert, insbesondere die starre Postulierung *der* allgemeingültigen Wissenschaftsmethodologie war Ansatz von Kritik. Nichtsdestoweniger bleibt es Poppers Verdienst, einen schlüssigen Rahmen für wissenschaftliches Vorgehen beschrieben zu haben, dem konzeptuell noch heute gefolgt wird (bspw. URHAHNE et al., im Druck). Auch vorliegendes Projekt entspricht dem deduktiven Ansatz, da aus einer Theorie heraus Items für die empirische Anwendung erarbeitet werden.

2.2 Modelle in der Wissenschaftsphilosophie

Die Bedeutung von Modellen für die Naturwissenschaft ist umstritten. Bis in die 1950er Jahre werden Modelle vornehmlich als ästhetische oder didaktische Hilfsmittel einer auf Theorien fußenden Naturwissenschaft (CARNAP, 1973, S.92) oder gar als „Schmarotzerpflanzen“ (DUHEM, 1908, S.37) naturwissenschaftlicher Theorien verstanden; als „models of known things and processes“ (HARRÉ, 1970, S.40; Herv. d. A.).

Diese Sichtweise wird gegen Mitte des 20. Jh. relativiert und die Relevanz von Modellen als Vehikel für die praktische Anwendung einer Theorie anerkannt (BAILER-JONES, 1999). Darüber hinaus wird Modellen eine bedeutende Rolle für die Konzeption neuer Theorien eingeräumt: Modelle können als Mittel einer „creative imagination“ (HESSE, 1953, S.198) dienen, welche die Theorieentwicklung erst ermöglicht. Aufgrund dieser Anwendungsmöglichkeiten können sie verstanden werden als „models for unknown mechanisms“ (HARRÉ 1970, S.40; Herv. d. A.).

MAHR (2004, 2008) zeigt, dass Modelle stets sowohl *models of known things and processes* als auch *models for unknown mechanisms* sind. Zwar sind nicht immer beide Orientierungen gleichermaßen stark ausgeprägt, sie sind aber entscheidend für die Beantwortung der Frage, ob es sich bei etwas um ein Modell handelt oder nicht (MAHR, 2008).

2.3 Modellkompetenz

Authentischer naturwissenschaftlicher Unterricht sollte folglich Modelle als Medium (*model of sth.*) und Methode (*model for sth.*) thematisieren (GILBERT, 2004). Diese Forderung findet sich auch in öffentlichen Zielvorgaben für allgemeinbildende Schulen. Die Bildungsstandards für das Fach Biologie formulieren im Kompetenzbereich Erkenntnisgewinnung das Anwenden von Modellen zur Veranschaulichung, Analyse, Beschreibung und Erklärung biologischer Aspekte (KMK, 2005). Gleichzeitig vertreten angehende LehrerInnen (CRAWFORD & CULLIN, 2004), erfahrene LehrerInnen (VAN DRIEL & VERLOOP, 2002) sowie SchülerInnen (GROSSLIGHT et al., 1991) eine vorwiegend naive¹ Sicht auf Modelle in den Naturwissenschaften. Unter LehrerInnen der naturwissenschaftlichen Fächer scheinen Biologielehrende diese naive Sicht in besonderem Maße zu besitzen (JUSTI & GILBERT, 2003).

Auf Grundlage der Studien über das Modellverständnis von LehrerInnen (JUSTI & GILBERT, 2003; CRAWFORD & CULLIN, 2004) und SchülerInnen (GROSSLIGHT et al., 1991) sowie unter Zuhilfenahme der Ergebnisse Mahrs (2008), haben UPMEIER ZU BELZEN & KRÜGER (2010) ein Kompetenzmodell der Modellkompetenz entwickelt (vgl. Tab. 1). Dieses beschreibt Modellkompetenz in den Dimensionen *Kenntnisse über Modelle* und *Modellbildung*. Die Dimension Kenntnisse über Modelle ist in zwei (*Eigenschaften von Modellen, Alternative Modelle*), die Dimension Modellbildung in drei Teilkompetenzen (*Zweck von Modellen, Testen von Modellen, Ändern von Modellen*) differenziert. Jede Teilkompetenz ist wiederum in drei Komplexitätsniveaus ausgeprägt.

Das Kompetenzmodell der Modellkompetenz wird aktuell im Rahmen mehrerer Promotionsvorhaben evaluiert. Ziel ist es, das Strukturmodell der Modellkompetenz zu validieren. Auf der Grundlage des validierten Kompetenzmodells sollen Diagnoseaufgaben für Modellkompetenz entwickelt werden. Dieses Vorhaben wird mit Hilfe offener Aufgaben (GRÜNKORN et al., 2009b), Multiple-Choice-Aufgaben (TERZER et al., 2009) und Forced-Choice-Aufgaben verfolgt. Im hier beschriebenen Projekt werden deduktiv Items aus dem Kompetenzmodell heraus entwickelt und statistisch analysiert. Die Ergebnisse fließen dann in die Entwicklung von Forced-Choice-Aufgaben.

¹ Naiv im Sinne einer (naiv-)realistischen Epistemologie (MATTHEWS, 1994).

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

Tab. 1: Kompetenzmodell der Modellkompetenz nach UPMEIER ZU BELZEN & KRÜGER (2010).

3 Fragestellungen

Für ihre Implementierung in Forced-Choice-Aufgaben werden für jedes Niveau jeder Teilkompetenz des Kompetenzmodells fünf Items formuliert. Die für die weitere Verwendung der Items zur Untersuchung des Kompetenzmodells grundsätzliche Fragestellung ist:

1. Inwieweit ist die dreigliedrige Struktur jeder Teilkompetenz des Kompetenzmodells durch deduktiv entwickelte Items abbildbar?

Insbesondere für die Itemselektion ist bedeutsam:

2. Welche drei Items eines Niveaus weisen die am stärksten übereinstimmenden Ergebnisse auf und sind daher geeignet, im weiteren Verlauf des Projekts Verwendung zu finden?

Für die weitere Planung des Projekts ist darüber hinaus interessant:

3. In welcher Weise stehen die Ergebnisse der Items unterschiedlicher Niveaus einer Teilkompetenz in einem Zusammenhang?

4 Methodik

Für jedes Niveau des Kompetenzmodells werden, in Anlehnung an den Fragebogen *My Views of Models and Modelling in Science* (TREGUST et al., 2004) und die Arbeiten von GRÜNKORN et al. (2009a), deduktiv Items formuliert und inhaltlich validiert (vgl. Tab. 2).

Die fünf Items eines Niveaus entsprechen inhaltlich dem, was theoretisch mit diesem Niveau beschrieben ist. Um die Synonymität der Items aus Sicht der ProbandInnen zu untersuchen und herauszufinden, welche Items eines Niveaus am stärksten miteinander korrelieren, wurden die Items in geschlossene Aufgaben mit einer 5-stufigen Ratingskala implementiert. ProbandInnen hatten die Aufgabe, ihre Zustimmung zu jedem Item im Kontext eines konkreten Modells zu gewichten. Die Skala zur Angabe der Zustimmungintensität wurde in Anlehnung an ROHRMANN (1978) äquidistant beschriftet (1: gar nicht – 2: kaum – 3: mittelmäßig – 4: überwiegend – 5: völlig; zusätzliche Option: weiß nicht). Es wurden insgesamt drei unterschiedliche Modelle eingesetzt, in deren Kontext die Items bewertet wurden: Ein Strukturmodell (Herzmodell), ein Funktionsmodell (Armmodell) und ein Denkmodell (Neandertalmodell).² Die Differenzierung in drei Modelltypen erfolgt nach GROPPENGIEBER & KATTMANN (2006).

Um die Bearbeitungslänge angemessen zu halten, hatten die ProbandInnen die Aufgabe, die Items lediglich dreier Teilkompetenzen im Kontext jeweils eines der drei Modelle zu bewerten. Um jede Kombination von Teilkompetenz und Modell zu realisieren und damit die Items jeder Teilkompetenz im Kontext jedes Modells bewerten zu lassen, wurden fünf Testhefte entworfen.

² Bei der Auswahl der Modelle hat die Erfahrung von JULIANE GRÜNKORN (FU Berlin) & EVA TERZER (HU Berlin) geholfen.

Niveau 1	Niveau 2	Niveau 3
Eigenschaften von Modellen		
[Das Modell] ...		
... gleicht [dem Original]. ... sieht so aus wie [das Original]. ... stimmt mit [dem Original] überein.	... zeigt zutreffend wichtige Bestandteile [des Originals]. ... macht wesentliche Merkmale [des Originals] sachgemäß deutlich. ... stellt sachgerecht bedeutsame Elemente [des Originals] dar.	... stellt dar, was man über [das Original] vermutet. ... bildet ab, welche Annahmen man über [das Original] hat. ... veranschaulicht, was man über [das Original] annimmt.
Alternative Modelle		
Es gibt verschiedene Modelle [des Originals] ...		
... weil sie sich in ihrem Aussehen unterscheiden. ... weil sie im Erscheinungsbild unterschiedlich sind. ... weil sie äußerlich verschieden sind.	... weil bei ihrer Herstellung unterschiedliche Aspekte [des Originals] einbezogen wurden. ... weil bei ihrer Herstellung unterschiedliche Eigenschaften [des Originals] berücksichtigt wurden. ... weil bei ihrer Herstellung unterschiedliche Elemente [des Originals] bedacht wurden.	... weil mit ihnen unterschiedliche Annahmen über [das Original] untersucht werden sollen. ... weil mit ihnen unterschiedliche Vermutungen über [das Original] erforscht werden sollen. ... weil mit ihnen unterschiedliche Hypothesen über [das Original] überprüft werden sollen.
Zweck von Modellen		
(Die Items der Teilkompetenz Zweck von Modellen sind in Tabelle 3 dargestellt.)		
Testen von Modellen		
[Das Modell] wird getestet ...		
... indem man [das Modell] auf Beschädigungen kontrolliert. ... indem man untersucht, ob es Schäden [am Modell] gibt.	... indem man seine Ähnlichkeit mit [dem Original] kontrolliert. ... indem man nachprüft, ob es [dem Original] entspricht. ... indem man untersucht, ob es [dem Original] ähnelt.	... indem man erforscht, ob neue Vermutungen über [das Original] zutreffen. ... indem man testet, ob neue Erwartungen an [das Original] zutreffen. ... indem man überprüft, ob neue Hypothesen über [das

... indem man nachsieht, ob [das Modell] Mängel besitzt.		Original] zutreffen.
Ändern von Modellen		
[Das Modell] wird verändert ...		
... um Beschädigungen [am Modell] zu beheben. ... um Schäden [am Modell] zu beseitigen. ... um Schwächen [am Modell] zu korrigieren.	... um neue Ergebnisse über [das Original] einzubeziehen. ... um neues Wissen über [das Original] einzubinden. ... um neue Einsichten über [das Original] zu beachten.	... um neue Annahmen über [das Original] untersuchen zu können. ... um neue Vermutungen über [das Original] erforschen zu können. ... um neue Hypothesen über [das Original] überprüfen zu können.

Tab. 2: Konstruierte Items. Die Platzhalter ([das Original] bzw. [das Modell]) werden in jeder Aufgabe durch das jeweilige Original oder Modell ersetzt. Aufgeführt sind lediglich die drei selektierten Items jedes Niveaus (vgl. Fragestellung 2).

Die Testhefte wurden sowohl bei Studierenden ($n > 450$), als ProbandInnen mit einem postulierten elaborierten Modellverständnis, als auch bei SchülerInnen ($n > 370$), als ProbandInnen mit einem postulierten weniger elaborierten Modellverständnis, eingesetzt. Es wurden Lehramtsstudierende der ersten vier Semester sowie SchülerInnen der Klassenstufen 7 bis 10 aus Gymnasien, Gesamtschulen und Realschulen befragt. Um die Synonymität der Items eines Niveaus innerhalb einer Teilkompetenz (und gleichzeitig die inhaltliche Abgrenzung gegenüber den Items der beiden anderen Niveaus derselben Teilkompetenz) zu testen, wurden die Ergebnisse faktorenanalytisch ausgewertet (vgl. Fragestellung 1). Um für jedes Niveau jeder Teilkompetenz die drei Items mit den am stärksten übereinstimmenden Ergebnissen zu identifizieren, wurde ergänzend die Reliabilität der Items eines Niveaus berechnet (vgl. Fragestellung 2). Einblicke in den Zusammenhang der Ergebnisse unterschiedlicher Niveaus einer Teilkompetenz wurden über eine Korrelationsanalyse gewonnen (vgl. Fragestellung 3). Sämtliche statistische Verfahren werden mit Hilfe von SPSS, Version 17, umgesetzt.

5 Ergebnisse

5.1 Ergebnisse der Faktorenanalyse

Für die Beantwortung der Fragestellungen 1 und 2 sind insbesondere die Ergebnisse der explorativen Faktorenanalyse interessant. Diese zeigen in allen Teilkompetenzen des Kompetenzmodells³ auf Anhieb drei Faktoren, auf denen die Items eines Niveaus separat laden. Im Folgenden werden exemplarisch die Ergebnisse der Teilkompetenzen *Zweck von Modellen* und *Eigenschaften von Modellen* präsentiert.⁴

Eine Hauptkomponentenanalyse mit orthogonaler Rotation (varimax) wurde mit 15 Items der Teilkompetenz *Zweck von Modellen* durchgeführt. Der KMO-Test bestätigt die Stichprobengröße ($KMO > 0,88$; „great“ nach FIELD, 2009). Alle MSA-Werte für die Einzelitems sind mit $> 0,84$ sehr gut. Der Bartlett-Test zeigt, dass die Korrelationen zwischen den Items ausreichend hoch sind. Drei Faktoren besitzen einen Eigenwert > 1 und erklären zusammen etwa 76% der Gesamtvarianz. Beides spricht für die Extraktion eben dieser Faktoren.

Auf den extrahierten Faktoren laden, jeweils spezifisch, die Items eines Niveaus der Teilkompetenz *Zweck von Modellen*. Sie können folglich inhaltlich als eben diese Niveaus verstanden werden. Tabelle 3 zeigt die Faktorladungen jener drei Items jedes Niveaus, die aufgrund statistischer Kennwerte und inhaltlicher Gesichtspunkte⁵ selektiert wurden. Die Subskalen für die Niveaus I, II und III, repräsentiert durch die drei dargestellten Items, weisen jeweils eine hohe Reliabilität aus (nach FIELD, 2009).

Die Items der Teilkompetenz *Eigenschaften von Modellen* lassen sich in den Niveaus I und II statistisch nicht befriedigend trennen. So laden Items von Niveau II sowohl auf demselben Faktor wie diejenigen von Niveau I, als auch auf einem weiteren Faktor. Die Items von Niveau III laden dagegen auf einem separaten Faktor. Aufgrund dieser Ergebnisse werden je drei Items des Niveaus I sowie drei Items des Niveaus III (vgl. Tab. 2) für die weitere Verwendung in Forced-Choice-Aufgaben ausgewählt.

³ Ausnahme: *Eigenschaften von Modellen*.

⁴ Alle weiteren Ergebnisse sind über die Autoren einsehbar.

⁵ Der Begriff Vorstellung hat sich in offenen Ansätzen (GRÜNKORN et al., 2009) als problematisch erwiesen, weshalb Items mit diesem Begriff nicht weiter verwendet werden.

		Faktor		
[Das Modell] hat den Zweck ...		1	2	3
I	... [das Original] darzustellen.	,904		
	... [das Original] zu veranschaulichen.	,891		
	... [das Original] anschaulich zu machen.	,884		
II	... Verbindungen zwischen Teilen [des Originals] begrifflich zu machen.		,870	
	... Verhältnisse zwischen Teilen [des Originals] verständlich zu machen.		,856	
	... Verknüpfungen zwischen Teilen [des Originals] zu erläutern.		,855	
III	... neue Vermutungen über [das Original] zu erforschen.			,910
	... neue Annahmen über [das Original] zu untersuchen.			,884
	... neue Hypothesen über [das Original] zu überprüfen.			,871
Reliabilität (Cronbachs α)		,875	,832	,873
Eigenwert		3,71	3,67	3,96
Erklärte Varianz [%]		24,74	24,47	26,39

Tab. 3: Ergebnisse der faktorenanalytischen Auswertung der Items für die Teilkompetenz *Zweck von Modellen*. N = 392 (SchülerInnen und Studierende), KMO = 0,88 (für Einzelitems MSA > 0,84), Bartlett-Test: $X^2(105) = 4232,60$ mit $p < 0,001$. Drei Faktoren mit einem Eigenwert > 1. Extraktionsmethode Hauptkomponentenanalyse; kleine Werte (< 0,3) unterdrückt. I, II, III: Niveaus im Kompetenzmodell.

Um die inhaltliche Nähe der Niveau II-Items zu denen von Niveau I zu reduzieren, wurden sie überarbeitet und gemeinsam mit den selektierten Items von Niveau I und III erneut getestet. Dieser zweite Test wurde bisher nur mit Studierenden (n = 76) durchgeführt⁶ und lieferte nun zufrieden stellende Ergebnisse: Der KMO-Test bestätigt die Stichprobengröße (KMO > 0,75; „good“ nach FIELD, 2009). Alle MSA-Werte für die Einzelitems sind mit > 0,5 akzeptabel. Der Bartlett-Test zeigt, dass die Korrelationen zwischen den Items ausreichend hoch sind. Drei Faktoren besitzen einen Eigenwert > 1 und erklären etwa 74% der Gesamtvarianz. Die Subskalen für die Niveaus I, II und III, repräsentiert durch drei Items (vgl. Tab. 2), weisen jeweils eine hohe Reliabilität aus (alle > 0,80, nach FIELD, 2009).

Aufgrund dieser Ergebnisse konnten auch in der Teilkompetenz *Eigenschaften von Modellen* drei Items in jedem Niveau für die weitere Verwendung in Forced-Choice-Aufgaben ausgewählt werden (vgl. Tab. 2). Die

⁶ Eine Untersuchung mit SchülerInnen läuft aktuell im Rahmen einer Staatsexamensarbeit.

Kennwerte der Items der übrigen Teilkompetenzen sind ähnlich unproblematisch wie diejenigen der Teilkompetenz *Zweck von Modellen*.

5.2 Ergebnisse der Korrelationsanalyse

Um einen ersten Einblick in einen möglichen Zusammenhang zwischen den Ergebnissen der Items der unterschiedlichen Niveaus einer Teilkompetenz zu erlangen (vgl. Fragestellung 3), wurden bivariate Korrelationen berechnet. Im Folgenden werden die Ergebnisse der Teilkompetenz *Zweck von Modellen* exemplarisch präsentiert (vgl. Tab. 4).⁷

Modelltyp		I	II	III
Strukturmodell	I (4,3)	1	0,355**	-0,149
	II (3,4)		1	0,194*
	III (2,2)			1
Funktionsmodell	I (3,3)	1	0,150*	0,258**
	II (3,9)		1	0,084
	III (2,8)			1
Denkmodell	I (4,2)	1	0,171*	0,084
	II (3,0)		1	0,538**
	III (3,0)			1

Tab. 4: Ergebnisse der Korrelationsanalyse (bivariat) für die Items der Teilkompetenz *Zweck von Modellen* im Kontext des Strukturmodells ($n = 163$), des Funktionsmodells ($n = 200$) und des Denkmodells ($n = 166$). I, II, III: Niveaus im Kompetenzmodell. In Klammern angegeben ist die mittlere Zustimmungintensität zu den Items des entsprechenden Niveaus. *: Korrelationen sind auf dem Niveau von 0,05 signifikant. **: Korrelationen sind auf dem Niveau von 0,01 signifikant.

Im Kontext des Strukturmodells und des Denkmodells sind die Korrelationen zwischen den Items von Niveau I und II bzw. von Niveau II und III signifikant positiv. Dabei ist beim Strukturmodell die Korrelation zwischen den Items von Niveau I und II, beim Denkmodell diejenige zwischen den Items von Niveau II und III am stärksten. Im Kontext des Funktionsmodells korrelieren die Ergebnisse der Items von Niveau I und II sowie von Niveau I und III signifikant positiv. Am stärksten korrelieren hier die Items von Niveau I und III.

⁷ Alle weiteren Ergebnisse sind über die Autoren einsehbar.

Die Unterschiede der mittleren Zustimmungswerte innerhalb der jeweiligen Modellkontexte sind, mit der Ausnahme von Niveau II und III im Kontext des Denkmodells, signifikant voneinander verschieden (nach dem t-Test). Im Kontext des Strukturmodells ist die mittlere Zustimmung zu den Items von Niveau III sehr niedrig (2,2). Diese unterscheidet sich außerdem signifikant von den Werten für die Items von Niveau III der beiden anderen Kontexte.

6 Diskussion

Das Vorgehen der vorliegenden Untersuchung ist dezidiert deduktiv und quantitativ ausgelegt. Die ausgewählten Items sind folglich primär aufgrund von statistischen Überlegungen selektiert worden. Die Items eines Niveaus können angesichts der faktorenanalytischen Ergebnisse zwar als einer inhaltlichen Dimension zugehörig betrachtet werden, die tatsächlichen inhaltlichen Überlegungen der ProbandInnen während der Aufgabenbearbeitung wurden bislang aber nicht untersucht. Da somit beispielsweise auch sprachliche Aspekte in die Bewertung der Items eingeflossen sein können, ist eine ergänzende qualitative Untersuchung sinnvoll.

Die Niveaus einer Teilkompetenz beschreiben ein inhaltliches Kontinuum. Das heißt, Niveau I und III sind inhaltlich am weitesten voneinander entfernt. Dieser Annahme entsprechen die starken Korrelationen zwischen den Ergebnissen der Items von Niveau I und II sowie Niveau II und III. Überraschend sind die Korrelationen im Kontext des Funktionsmodells. Hier korrelieren die Ergebnisse der Items von Niveau I und III am stärksten. Dies könnte ein Hinweis darauf sein, dass es sich bei den Niveaus des Kompetenzmodells nicht um Entwicklungsstufen handelt, sondern vielmehr um unterschiedliche Qualitäten, die relativ unabhängig voneinander sind. Ein qualitativer Zugang kann konkreter Einblicke auch in diesen Zusammenhang gewähren.

Die mittlere Zustimmung für die Items von Niveau III ist im Kontext des Strukturmodells sehr niedrig. Für weitere Untersuchungen werden daher keine Strukturmodelle mehr verwendet, da diese offenbar nur relativ geringe Zustimmung zu den komplexeren Niveaus II und vor allem III provozieren. Diese Beobachtung ist aus theoretischer Sicht damit zu begründen, dass Strukturmodelle in der Regel als abbildendes Medium der Wirklichkeit eingesetzt werden. Eine Kontextabhängigkeit des Modellverständnisses wurde

bereits in der Interviewstudie von TRIER & UPMEIER ZU BELZEN (2009) beschrieben. Dort wurde gezeigt, dass das Modellverständnis von SchülerInnen auch davon abhängt, in welcher Weise sie ein Modell bisher, also vermutlich im Biologieunterricht, erlebt haben.

Aufgrund dieser Einblicke scheint es sinnvoll, weitere Untersuchungen mit Forced-Choice-Aufgaben durchzuführen. Dieser Aufgabentyp erzeugt ipsative, das heißt relative Daten (HICKS, 1970) und erlaubt damit, auch bei sehr ähnlich starken, also potentiell uneindeutigen Zustimmungen, intrapersonale Präferenzunterschiede zu ermitteln (CHRISTIANSEN, BURNS & MONTGOMERY, 2005). So können etwaige Probleme, wie sie beispielsweise durch die Verwendung von klassischen Multiple-Choice- oder Ratingskalen-Aufgaben auftreten können, vermieden werden. Denn SchülerInnen, die mehreren Niveaus sehr ähnlich stark zustimmen, würden mit diesen Aufgaben gegebenenfalls nur schwierig interpretierbare Ergebnisse produzieren. Forced-Choice-Aufgabe hingegen zwingen ProbandInnen dazu, eine Präferenz anzugeben. Sie ergeben somit auch dann eine Präferenzrangordnung, wenn die Unterschiede in der Zustimmungintensität zu gering sind, als dass sie mit anderen Aufgabenformaten erfassbar wären.

7 Ausblick

Durch das skizzierte Vorgehen sind für jede Zelle des Kompetenzmodells drei Items konstruiert worden, die sowohl untereinander sehr ähnliche Bewertungen provoziert haben als auch von den Items der anderen Niveaus (aber innerhalb derselben Teilkompetenz) statistisch getrennt werden konnten. Für die weitere Evaluation des Kompetenzmodells und die Konstruktion eines Diagnoseinstruments werden die Items sowohl in Forced-Choice-Aufgaben als auch in Ratingskalen-Aufgaben implementiert. Forced-Choice-Aufgaben ergeben ipsative Daten, während Ratingskalen-Aufgaben normative Daten erzeugen.

Beide Aufgabentypen werden im Sommer 2010 gemeinsam mit offenen Aufgaben (GRÜNKORN et al., 2009b) eingesetzt, um über einen Vergleich einen Einblick in die Güte der unterschiedlichen Messkonstrukte zu erlangen (Konstruktvalidierung).

Die Entscheidungsprozesse von SchülerInnen werden beim Bearbeiten der Forced-Choice-Aufgaben und Ratingskalen-Aufgaben durch Lautes Denken (ERICSSON & SIMON, 1984) zusätzlich qualitativ überprüft. Hierdurch lassen sich die Aufgaben inhaltlich aus Sicht der ProbandInnen validieren. Außerdem

können unter Umständen Rückschlüsse auf die im Kompetenzmodell postulierte Struktur von Modellkompetenz und Einblicke in die Abhängigkeit von Modellkompetenz vom Kontext gewonnen werden.

Voraussichtlich Anfang 2011 werden die auf diesem Wege validierten Aufgaben im Rahmen einer gemeinsamen Hauptuntersuchung zusammen mit offenen Aufgaben (GRÜNKORN et al., 2009b) und Multiple-Choice-Aufgaben (TERZER et al., 2009) eingesetzt, um die Ergebnisse zwecks einer wechselseitigen (Konstrukt-)Validierung vergleichen zu können. Darüber hinaus werden parallel allgemeine Fähigkeiten (Lesekompetenz, Intelligenz) als Kontrollvariablen erhoben.

Vorliegendes Projekt ist damit Teil einer Kooperation mehrerer Ansätze, die darauf abzielen, auf Grundlage des Kompetenzmodells der Modellkompetenz Diagnoseaufgaben für den schulischen Einsatz zu entwickeln. Abgrenzende Charakteristika des vorliegenden Projekts sind sowohl das explizit theoriegeleitete (deduktive) Vorgehen als auch die Verwendung von Forced-Choice-Aufgaben.

Zitierte Literatur

- BAILER-JONES, D.M. (1999): Tracing the Development of Models in the Philosophy of Science. In: MAGNANI, L., N. J. NERSESSIAN & P. THAGARD [Hrsg.]: *Model-Based Reasoning in Scientific Discovery*. Kluwer Academic/Plenum Publishers, New York, 23-40.
- CARNAP, R. (1973): *Grundlagen der Logik und Mathematik*. Nymphenburger Verlagshandlung, München.
- CRAWFORD, B. A. & M. J. CULLIN (2004): Supporting Prospective Teachers' Conceptions of Modeling in Science. *International Journal of Science Education* **26**, 11, 1379-1401.
- CHRISTIANSEN, N.D., G.N. BURNS & G.E. MONTGOMERY (2005): Reconsidering Forced-Choice Item Formats for Applicant Personality Assessment. *Human Performance* **18**, 3, 267-307.
- DUHEM, P. (1908): *Ziel und Struktur der physikalischen Theorie*. Verlag von Johann Ambrosius Barth, Leipzig.
- ERICSSON, K. A. & H.A. SIMON (1984): *Protocol Analysis. Verbal Reports as Data*. The MIT Press: Cambridge & London.
- FIELD, A. (2009): *Discovering Statistics Using SPSS. Third Edition*. Sage Publications Ltd., Los Angeles u.a.
- GILBERT, J.K. (2004): Models and Modelling: Routes to More Authentic Science Education. *International Journal of Science and Mathematics Education* **2**, 115-130.
- GILBERT, S.W. (1991): Model Building and a Definition of Science. *Journal of Research in Science Teaching* **28**, 1, 73-79.
- GROPENGIEBER, H. & U. KATTMANN (2006): *Fachdidaktik Biologie. 7., völlig überarbeitete Auflage*. Aulis, Köln.
- GROSSLIGHT, L., C. UNGER, E. JAY & C.L. SMITH (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., A. UPMEIER ZU BELZEN & D. KRÜGER (2009a): Modellkompetenz im Biologieunterricht – Überprüfung eines Kompetenzmodells mit offenen und geschlossenen Aufgaben. In: Referierter Tagungsband der 11. Internationalen Frühjahrsschule der Fachsektion Didaktik der Biologie (FDdB) im VBIO, Universität Gießen, 05.-08.03.2009, 18-19.
- GRÜNKORN, J., A. UPMEIER ZU BELZEN & D. KRÜGER (2009b): Diagnose von Modellkompetenz im Biologieunterricht – Validierung eines Kompetenzmodells mit offenen und halboffenen Aufgaben. In: Referierter Tagungsband der Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, Universität Kiel, 21.-25.09.2009, 172-174.
- HARRÉ, R. (1970): *The Principles of Scientific Thinking*. Macmillan and Co Ltd., London & Basingstoke.
- HESSE, M.B. (1953): Models in Physics. *The British Journal for the Philosophy of Science* **4**, 15, 198-214.
- HICKS, L.E. (1970): Some Properties of Ipsative, Normative, and Forced-Choice Normative Measures. *Psychological Bulletin* **74**, 3, 167-184.
- JUSTI, R. & J. GILBERT (2003): Teacher's Views on the Nature of Models. *International Journal of Science Education* **25**, 1, 1369-1386.
- SEKRETARIAT DER STÄNDIGEN KONFERENZ DER KULTUSMINISTER DER LÄNDER IN DER BUNDESREPUBLIK DEUTSCHLAND (KMK) (2005): *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. Wolters Kluwer, München & Neuwied.
- MAHR, B. (2004): *Das Wissen im Modell*. KIT-Report 150. TU-Berlin, 1-21.
- MAHR, B. (2008): Ein Modell des Modellseins. Ein Beitrag zur Aufklärung des Modellbegriffs. In: DIRKS, U. & E. KNOBLOCH [Hrsg.]: *Modelle*. Peter Lang, Frankfurt am Main, 187-218.
- MATTHEWS, M. R. (1994): *Science Teaching. The Role of History and Philosophy of Science*. Routledge, New York & London.
- POPPER, K. (1934/2005): *Logik der Forschung*. In: KEUTH, H. [Hrsg.]: *Karl. R. Popper. Gesammelte Werke in deutscher Sprache 3*. Mohr Siebeck, Tübingen.
- POPPER, K. (1963/2009): *Vermutungen und Widerlegungen. Das Wachstum der wissenschaftlichen Erkenntnis*. In: KEUTH, H. [Hrsg.]: *Karl. R. Popper. Gesammelte Werke in deutscher Sprache 10*. Mohr Siebeck, Tübingen.
- ROHRMANN, B. (1978): Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie* **9**, 222-245.
- TERZER, E., D. KRÜGER & A. UPMEIER ZU BELZEN (2009): Diagnose von Modellkompetenz im Biologieunterricht - Validierung eines Kompetenzmodells mit Multiple Choice-Aufgaben. Referierter Tagungsband der Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, Universität Kiel, 21.-25.09.2009, 140-141.
- TERZER, E. & A. UPMEIER ZU BELZEN (2007): Naturwissenschaftliche Erkenntnisgewinnung durch Modelle – Modellverständnis als Grundlage für Modellkompetenz. *Berichte aus Institutionen der Didaktik der Biologie* **16**, 33-56.
- TREAGUST, D.F., G. CHITTLEBOROUGH & T.L. MAMIALA (2004): Students' Understanding of the Descriptive and Predictive Nature of Teaching Models in Organic Chemistry. *Research in Science Education* **34**, 1-20.
- TRIER, U. & A. UPMEIER ZU BELZEN (2009): Schülervorstellungen zu Modellen als Basis für die Entwicklung von Modellkompetenz. Referierter Tagungsband der Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, Universität Kiel, 21.-25.09.2009, 74-75.
- UPMEIER ZU BELZEN, A. & D. KRÜGER (2010): Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften* **16**, 59-75.
- URHAHNE, D., S. MARSCH, M. WILDE & D. KRÜGER (im Druck): Die Messung konstruktivistischen Unterrichts. *Psychologie in Erziehung und Unterricht*.

VAN DRIEL, J. H. & N. VERLOOP (2002): Experienced Teacher's Knowledge of Teaching and Learning of Models and Modeling in Science Education. *International Journal of Science Education* **24**, 12, 1255-1277.

