

## **Forced Choice-Aufgaben zur Evaluation von Modellkompetenz im Biologieunterricht**

### **Empirische Überprüfung konstrukt- und merkmalsbezogener Teilkompetenzen**

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**

*Modelle und das Modellieren sind wichtige Bestandteile naturwissenschaftlichen Unterrichts. Dieser Beitrag beschreibt die Entwicklung von Forced Choice-Aufgaben zur empirischen Evaluation eines Kompetenzmodells für Modellkompetenz im Biologieunterricht. Die im Artikel diskutierten Fragestellungen beziehen sich auf die Überprüfung konstrukt- sowie inhaltsbezogener Teilkompetenzen des Kompetenzmodells der Modellkompetenz. Beide Fragen werden auf der Grundlage von Ergebnissen einer empirischen Studie untersucht (7. - 10. Klasse, N = 901). Die Befunde sprechen für die theoretisch in Teilkompetenzen beschriebene, fünfdimensionale Struktur des Kompetenzmodells. Die Prüfung inhaltsbezogener Teilkompetenzen fällt weniger eindeutig aus, die Daten sprechen aber gegen den Inhalt als schwierigkeiterzeugendes Aufgabenmerkmal.*

#### **Abstract**

*Models and modelling are important elements of science education. The present article describes the design of forced choice tasks and their application for the empirical evaluation of a model of model competence in biology education. The discussed issues refer to the empirical evaluation of construct-related as well as content-related dimensions of the model of model competence. Both points are discussed on the basis of the results of an empirical survey with 7th to 10th graders (N = 901). The results clearly support the five dimensional structure of model competence, as it is described theoretically. The results of the evaluation of content-related dimensions are less clear. However, the task-difficulty does not seem to be affected by the task-content.*

## 1 Einleitung

Die Ziele der allgemeinbildenden Schule werden in Deutschland in Form von Kompetenzen beschrieben. So umfassen bspw. die Bildungsstandards für den Mittleren Schulabschluss im Fach Biologie die Kompetenzbereiche Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung (KMK, 2005). Mit der Einführung der Bildungsstandards wurde bemängelt, dass „noch keine empirisch abgesicherten Kompetenzstufenmodelle vorliegen“ (a. a. O.: 16), weshalb sich die Formulierung unterschiedlicher Anforderungsbereiche der Bildungsstandards an den Einheitlichen Prüfungsanforderungen in der Abiturprüfung Biologie (EPA) orientiert hat (KMK, 1989/2004). Mittlerweile liegen für viele Bereiche der Biologie eben solche Modelle vor. Für Modellkompetenz im Biologieunterricht existiert ein theoretisch begründetes, aber nicht empirisch abgesichertes Kompetenzmodell (UPMEIER ZU BELZEN & KRÜGER, 2010). Die drei Anforderungsbereiche der EPA unterscheiden sich im Grad der Selbstständigkeit bei der Aufgabenbearbeitung sowie im Grad der Komplexität der zur Aufgabenbearbeitung notwendigen kognitiven Prozesse (KMK, 1989/2004: 14). Diese beiden Aspekte können als schwierigkeitserzeugende Aufgabenmerkmale bezeichnet werden. Neben diesen, den EPA zu Grunde liegenden Merkmalen, werden in der Literatur noch weitere Merkmale dahingehend diskutiert, inwieweit sie sich messbar auf die Aufgabenschwierigkeit auswirken. Hierzu gehören z. B. inhaltliche Aspekte oder auch das jeweilige Aufgabenformat (vgl. SCHMIEMANN, 2010). Diese Merkmale werden bei der Formulierung von Kompetenzmodellen in Form von *merkmalsbezogenen Teilkompetenzen* (PRENZEL et al., 2002) berücksichtigt, wenn davon ausgegangen wird (theoretisch) bzw. gezeigt werden kann (empirisch), dass sie sich messbar auf die Schwierigkeit der Aufgaben auswirken. In Abgrenzung hierzu können die auf das jeweilige Kompetenzkonstrukt bezogenen Teilkompetenzen als *konstruktbezogene Teilkompetenzen* bezeichnet werden. Je nach Forschungsprojekt existieren unterschiedliche Modelle für die Beschreibung und Kategorisierung schwierigkeitserzeugender Aufgabenmerkmale. PRENZEL et al. (2002) betonen, dass sich empirische Ergebnisse zu dem Einfluss einzelner Aufgabenmerkmale nicht ohne Weiteres generalisieren lassen. Daraus folgt, dass der Einfluss einzelner schwierigkeitserzeugender Aufgabenmerkmale bei der Formulierung von Kompetenzmodellen projektspezifisch untersucht und beschrieben werden sollte.

Vorliegender Artikel beschreibt die empirische Überprüfung eines Kompetenzmodells zur Modellkompetenz im Biologieunterricht mit Forced Choice-

Aufgaben. Hierbei wird zunächst die theoretisch postulierte Dimensionalität des Kompetenzmodells überprüft (Überprüfung konstruktbezogener Teilkompetenzen). Anschließend werden mögliche schwierigkeiterzeugende Aufgabenmerkmale diskutiert und der Einfluss des fachlichen Inhalts auf die Schülerantworten empirisch untersucht (Überprüfung merkmalsbezogener Teilkompetenzen).

## 2 Theorie

### 2.1 Forced Choice-Aufgaben

Forced Choice (FC)-Aufgaben sind geschlossene Aufgaben, die von den Probanden eine Bewertung gegebener Antwortalternativen ohne vorgegebene Skalenpunkte verlangen. Ein Beispiel für FC-Aufgaben sind Rangordnungsaufgaben, bei denen die präsentierten Antwortalternativen direkt in eine Rangfolge gebracht werden müssen (Abb. 1). Daher erzeugen FC-Aufgaben relative Urteile bzw. „ipsative Daten“ (HICKS, 1970).

Abbildung 1 zeigt das **Modell eines Palmen-Blattes**.

Palmen-Blätter können sehr unterschiedliche Formen und Strukturen haben. Abbildung 2 zeigt ein **Palmen-Blatt** mit einer gefalteten Blattfläche.

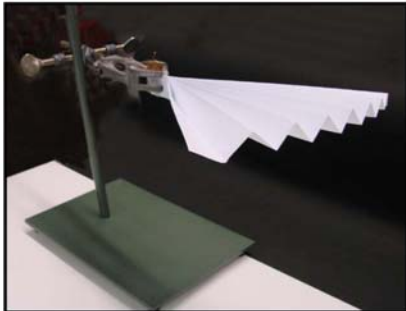


Abb. 1: Modell eines Palmen-Blattes




Abb. 2: Palmen-Blatt

**Gib an, welchen Zweck das Palmen-Blatt-Modell für Wissenschaftler hat!**

Wie stark treffen die folgenden Aussagen zu?  
Schreibe die Buchstaben **A** (trifft am stärksten zu), **B** (trifft mittelmäßig zu) oder **C** (trifft am wenigsten zu) neben jede Aussage.

Das Palmen-Blatt-Modell hat den Zweck, ...	
... das Palmen-Blatt möglichst genau darzustellen.	<input type="checkbox"/>
... Zusammenhänge zwischen Teilen des Palmen-Blattes zu verdeutlichen.	<input type="checkbox"/>
... mit dem Modell Vermutungen über das Palmen-Blatt zu gewinnen.	<input type="checkbox"/>

**Abb. 1:** Palmen-Aufgabe für die Teilkompetenz Zweck von Modellen. Die Antwortalternativen repräsentieren die Niveaus I. II. III (von oben nach unten).

Die direkt aus den Rangordnungs-Aufgaben gewonnenen Daten können nach CHAN (2003) als ordinale ipsative Daten bezeichnet werden. Werden nicht alle Informationen in die Datenanalyse einbezogen, wird bspw. nur die erste Präferenz ausgewertet, bringen FC-Aufgaben „partially ipsative data“ (HICKS, 1970) hervor. Im Folgenden wird für die Datenauswertung eben dieses Vorgehen gewählt, da gezeigt werden konnte (KRELL et al., *eing.*), dass bei den FC-Aufgaben zur Modellkompetenz im Biologieunterricht in einigen Fällen nur die erste Präferenz inhaltlich valide ausfällt.

## 2.2 Konstruktbezogene Teilkompetenzen

Das auf der Grundlage empirischer Studien über Schülervorstellungen zu Modellen (z. B. GROSSLIGHT et al., 1991) und theoretischer Erörterungen zum Modellbegriff (MAHR, 2008) erarbeitete Kompetenzmodell der Modellkompetenz (UPMEIER ZU BELZEN & KRÜGER, 2010) strukturiert Modellkompetenz und beschreibt somit die Voraussetzungen für einen kompetenten Umgang mit Modellen. Modellkompetenz ist nach UPMEIER ZU BELZEN und KRÜGER (2010) in die beiden Dimensionen Kenntnisse über Modelle und Modellbildung differenziert. Die erste Dimension umfasst die beiden Teilkompetenzen Eigenschaften von Modellen und Alternative Modelle, die zweite Dimension die Teilkompetenzen Zweck, Testen und Ändern von Modellen. Jede Teilkompetenz kann in jeweils drei unterschiedlich elaborierten Qualitäten<sup>2</sup> ausgeprägt sein. Das theoretische Konstrukt Modellkompetenz im Biologieunterricht besitzt nach UPMEIER ZU BELZEN und KRÜGER (2010) demnach fünf konstruktbezogene Teilkompetenzen. Diese Struktur gilt es empirisch zu überprüfen.

## 2.3 Merkmalsbezogene Teilkompetenzen

Eine nur auf statistischen Analysen beruhende Aufgabenkonstruktion und -selektion, in deren Zuge Aufgaben mit schlechten Kennwerten aussortiert werden, kann zu einer künstlichen Varianzreduktion führen und somit die Generalisierbarkeit von Ergebnissen in Frage stellen (BLUM et al., 2001, GOLDSTEIN, 2004). PRENZEL et al. (2002) betonen, dass die Analyse schwierigkeiterzeugender Aufgabenmerkmale wichtig ist, um die Existenz „merkmals-

---

<sup>2</sup> UPMEIER ZU BELZEN und KRÜGER sprechen von drei „Niveaus“, deren „Ordnung sowie eine potenzielle Entwicklung“ es empirisch zu überprüfen gilt (2010: 52).

bezogener Teilkompetenzen“ (2002: 124) zu überprüfen. Im Zusammenhang mit Modellkompetenz im Biologieunterricht sollte somit überprüft werden, inwiefern neben den konstruktbezogenen auch merkmalsbezogene Teilkompetenzen beschrieben werden können.

Allgemein können Merkmale von Aufgaben, die sich auf das Antwortverhalten auswirken, als schwierigkeiterzeugende Aufgabenmerkmale bezeichnet werden. Es existieren unterschiedliche Vorschläge zur Systematisierung dieser Merkmale (z. B. PRENZEL et al., 2002, ROST et al., 2005, SENKBEIL et al., 2005, SCHECKER & PARCHMANN, 2006, KAUERTZ, 2008, KAUERTZ et al., 2010). Aufbauend auf den Ergebnissen von PRENZEL et al. (2002) unterscheidet KAUERTZ (2008) zwischen formalen und inhaltlichen Aufgabenmerkmalen sowie kognitiven Verarbeitungsprozessen, wobei die beiden ersten als objektive und das letztere als prozessorientiertes Merkmal bezeichnet werden (2008: 22). Zu den objektiven Merkmalen werden gezählt: das Antwortformat, die Repräsentationsform, die Textlänge sowie inhaltliche Merkmale (KAUERTZ, 2008, SCHMIEMANN, 2010); bei SCHECKER und PARCHMANN (2006) auch der Aufgabenkontext.<sup>3</sup> Studien (GROSSLIGHT et al., 1991, KRELL & KRÜGER, 2010, GRÜNKORN et al., angeh.) legen außerdem nahe, dass bei Fragen zur Modellkompetenz das Antwortverhalten der Schüler vom präsentierten Modell beeinflusst werden kann. Prozessorientierte Merkmale beziehen sich auf die zum Lösen der Aufgaben erforderlichen kognitiven Prozesse und lassen sich nach KAUERTZ et al. (2010) dahingehend unterscheiden, ob sie auf die Beziehung zum Vorwissen, fachübergreifende Prozesse oder generelle Informationsverarbeitungsstrategien fokussieren. Im Folgenden wird auf den Inhalt als potenziell schwierigkeiterzeugenden Faktor fokussiert, da der fachliche Inhalt eine „aus fachdidaktischer Sicht besonders relevante Kategorie zur Charakterisierung von Aufgaben“ ist (SCHMIEMANN, 2010: 28). Bezogen auf die innerfachliche Perspektive der Biologie ist eine historisch-genetische Differenzierung des Inhalts (NEUMANN et al., 2007) in verschiedene biologische Disziplinen möglich. Da Modellkompetenz dem Kompetenzbereich Erkenntnisgewinnung zuzuordnen ist und „inhaltliche Merkmale [...] oft nicht mit Schwierigkeit in Zusammenhang gebracht werden“ (SENKBEIL, 2008: 23), wird bei der Untersuchung des Einflusses des Fachinhalts ein konfirmatorischer Ansatz verfolgt: Es werden je zwei Aufgaben aus den Disziplinen Zoologie, Botanik und Humanbiologie eingesetzt. Der Inhaltsaspekt wird im Sinne einer

---

<sup>3</sup> SCHECKER und PARCHMANN (2006) differenzieren zwischen drei Kontexten: innerfachlich, persönlich-gesellschaftliches Umfeld und professionelle Anwendungen. GILBERT (2006) weist allerdings darauf hin, dass der Kontextbegriff nicht einheitlich definiert ist.

Kontrollvariablen genutzt, um die Existenz merkmalsbezogener Teilkompetenzen zu prüfen.

### 3 Fragestellungen

Das Ziel dieses Beitrags ist eine Dimensionalitätsprüfung der Schülerantworten bei FC-Aufgaben zur Modellkompetenz im Biologieunterricht. Hierbei werden sowohl konstruktbezogene als auch merkmalsbezogene Teilkompetenzen untersucht. Die erste Fragestellung bezieht sich auf die empirische Überprüfung konstruktbezogener Teilkompetenzen:

**Fragestellung 1:** Inwiefern zeigt sich die zwei oder fünfdimensionale Struktur des Kompetenzmodells der Modellkompetenz empirisch in den Ergebnissen der FC-Aufgaben?

**Hypothese 1:** Es wird davon ausgegangen, dass sich die zweidimensionale Struktur nach den Dimensionen, ggf. aber auch die fünfdimensionale Struktur nach den theoretisch beschriebenen Teilkompetenzen des Kompetenzmodells der Modellkompetenz, empirisch zeigt (UPMEIER ZU BELZEN & KRÜGER, 2010).

Die zweite Fragestellung bezieht sich auf die Überprüfung merkmalsbezogener (hier: inhaltsbezogener) Teilkompetenzen:

**Fragestellung 2:** Inwieweit ist der Inhalt der Aufgaben ein Prädiktor für deren Schwierigkeit bzw. lassen sich inhaltsbezogene Teilkompetenzen empirisch nachweisen?

**Hypothese 2:** Der fachliche Inhalt zeigt sich nur selten als schwierigkeits-erzeugend (SENKBEIL, 2008). Es wird daher vermutet, dass die Ergebnisse der FC-Aufgaben ungeachtet der unterschiedlichen fachlichen Inhalte übereinstimmend ausfallen. D. h. es sollten sich keine inhaltsbezogenen Teilkompetenzen zeigen.

## 4 Methodik

### 4.1 Entwicklung der Forced Choice-Aufgaben

KRELL & KRÜGER (2010) haben auf der Grundlage des Kompetenzmodells der Modellkompetenz Items formuliert, statistisch analysiert und selektiert. Aus

diesen wurden FC-Aufgaben konstruiert<sup>4</sup> und inhaltlich validiert (KRELL et al., einger.). Das Gerüst der Aufgaben bilden allgemeine Formulierungen, die eine Operationalisierung des Kompetenzmodells der Modellkompetenz darstellen (Tab. 1). Diese Konstruktion erlaubt den Einsatz unterschiedlicher Aufgabenstämme, wodurch eine systematische Untersuchung einzelner schwierigkeits-erzeugender Aufgabenmerkmale möglich ist. Im Sinne eines Facettendesigns (MARTINEZ, 1999) können somit gezielt einzelne Merkmale des Aufgabenstamms (z. B. der Inhalt) variiert werden. Dies entspricht der Empfehlung SCHECKERS und PARCHMANN'S (2006), für die Überprüfung merkmalsbezogener Teilkompetenzen „nur jeweils Ausschnitte der Struktur“ zu untersuchen „und dabei andere Bereiche des Gesamtmodells zu kontrollieren“ (2006: 60).

**Tab. 2:** Operationalisierung des Kompetenzmodells der Modellkompetenz (vgl. KRELL et al., einger.). Die Platzhalter [] werden in den Aufgaben durch das jeweilige Modell bzw. Original ersetzt. I, II, III: Niveaus im Kompetenzmodell.

<b>Teilkompetenz Eigenschaften von Modellen</b>	
	[Das Modell] ...
I	... sieht so aus wie [das Original].
II	... zeigt zutreffend wichtige Merkmale [des Originals].
III	... stellt dar, was man über [das Original] vermutet.
<b>Teilkompetenz Alternative Modelle</b>	
	Es gibt verschiedene Modelle [des Originals], ...
I	... weil sich die Modelle äußerlich unterscheiden.
II	... weil bei der Herstellung der Modelle unterschiedliche Eigenschaften [des Originals] berücksichtigt wurden.
III	... weil mit den Modellen unterschiedliche Vermutungen über [das Original] gewonnen werden sollen.
<b>Teilkompetenz Zweck von Modellen</b>	
(Die Formulierungen dieser Teilkompetenz sind Abb. 1 zu entnehmen.)	
<b>Teilkompetenz Testen von Modellen</b>	
	[Das Modell] wird getestet ...
I	... indem man überprüft, ob [das Modell] beschädigt ist.
II	... indem man überprüft, ob [das Modell] [dem Original] entspricht.
III	... indem man mit [dem Modell] überprüft, ob Vermutungen über [das Original] zu treffen.

<sup>4</sup> Bei der Auswahl der Aufgabenstämme hat die Erfahrung von E. Terzer (HU Berlin) und J. Grünkorn (FU Berlin) geholfen.

<b>Teilkompetenz Ändern von Modellen</b>	
	[Das Modell] wird verändert ...
I	... um zu beheben, was am [Modell] beschädigt ist.
II	... um in [das Modell] neue Erkenntnisse über [das Original] einzubeziehen.
III	... um mit [dem Modell] weitere Vermutungen über [das Original] überprüfen zu können.

## 4.2 Fragebogenerhebung und Auswertung der Daten

Es wurden sechs unterschiedliche Aufgabenstämme in jeder Teilkompetenz des Kompetenzmodells eingesetzt, insgesamt also 30 Aufgaben. Um einen Vergleich zwischen den Teilkompetenzen möglichst unverzerrt zu erlauben, wurden dieselben Aufgabenstämme übereinstimmend in allen Teilkompetenzen verwendet (Facettendesign mit stammäquivalenten Aufgaben; TRAUB & MACRURY, 1990). In den sechs Aufgabenstämmen wird jeweils ein gegenständliches Modell beschrieben, das sich auf einen Struktur-Funktions-Zusammenhang bezieht. Die Modelle sind somit als Struktur und als Funktionsmodell interpretierbar (vgl. exemplarisch Abb. 1). Sie sind fachlich der Humanbiologie, der Botanik bzw. der Zoologie zuzuordnen (Tab. 2). In jeder Aufgabe müssen die Probanden die drei Niveaus einer Teilkompetenz in eine Präferenzrangfolge bringen (Abb. 1). Um den Arbeitsaufwand für die Schüler (N = 901, Jgst. 7-10, Gymnasium und Realschule) in angemessenen Grenzen zu halten, wurden mehrere Testhefte konzipiert, die jeweils nur einen Teil der 30 Aufgaben enthalten (Block-Design; von Davier, Carstensen & von Davier, 2008).

**Tab. 3:** Modelltyp, Inhalt und Modell der in allen Teilkompetenzen eingesetzten FC-Aufgaben.

<b>Modelltyp</b>	<b>Inhalt</b>	<b>Modell</b>
Struktur-Funktions-Zusammenhang	Humanbiologie	Mund
		Fuß
	Botanik	Samen
		Palme
	Zoologie	Libelle
		Vogel

Für die Beantwortung der Fragestellung 1 wurden die Daten mit Hilfe des Programms ConQuest (WU et al., 2007) auf Passung mit einem ein-, einem zwei- und einem fünfdimensionalen Partial Credit-Modell (MASTERS, 1982) geprüft. Das eindimensionale Modell subsumiert die Ergebnisse aller Aufgaben unter eine Dimension, unterstellt damit eine generelle, von den fünf Teilkom-



petenzen unabhängige Modellkompetenz und damit das Fehlen konstruktbezogener Teilkompetenzen. Das zweidimensionale Modell unterscheidet die beiden im Kompetenzmodell beschriebenen Dimensionen Kenntnisse über Modelle und Modellbildung, während das fünfdimensionale Modell schließlich die fünf Teilkompetenzen als einzelne Dimensionen betrachtet. Es wurden somit, neben dem generellen, eindimensionalen Modell, die beiden theoretisch zu erwartenden Modelle gewählt („choose the model first“; YEN & FITZPATRICK, 2006). Die drei Modelle wurden mittels AIC und BIC verglichen (BURNHAM & ANDERSON, 2004). Hierzu wurden neben den absoluten Werten (AIC und BIC) auch die durch einen Modellvergleich errechneten Differenzen  $\Delta AIC_i$  bzw.  $\Delta BIC_i$  sowie deren normalisierte Form *Akaike-Weights* ( $\omega_i$ ) bzw. *posterior model probabilities* ( $p_i$ ) berechnet. Letztere geben die Wahrscheinlichkeit an, mit der ein Modell das quasi-wahre Modell ist (BURNHAM & ANDERSON, 2004).

Für die Beantwortung der Fragestellung 2 wurden die Daten für einen ersten Überblick mit Hilfe von Kreuztabellen und dem auf diesen beruhenden Kennwert Cramer's V geprüft. Dieser erlaubt eine Einschätzung der Stärke des Zusammenhangs verschiedener Merkmalskombinationen, ist auch auf kategoriale, nicht normal verteilte Daten anwendbar (FIELD, 2009) und analog der parametrischen Produkt-Moment-Korrelation interpretierbar (BORTZ et al., 2008). Darüber hinaus wurden die Daten mit ConQuest dahingehend analysiert, ob inhaltsbezogene Teilkompetenzen vorliegen. Hierzu wurde ein Modell mit drei Dimensionen für die drei Inhaltsbereiche spezifiziert. Die Güte dieses Modells wurde äquivalent zur oben dargelegten Weise analysiert. Schließlich wurden die generellen Itemparameter des quasi-wahren Modells verwendet, um den Inhalt als möglichen Prädiktor der Aufgabenschwierigkeit in einer Regressionsanalyse zu prüfen (vgl. PRENZEL et al., 2002). Hierzu wurden die drei Inhaltsbereiche (Tab. 2) als Prädiktoren in einer linearen Regressionsanalyse geprüft (jeweils kodiert als dichotome Dummy-Variable).

## 5 Ergebnisse

### 5.1 Fragestellung 1: Prüfung konstruktbezogener Teilkompetenzen

Die Kennwerte des Modellvergleichs sprechen für das fünfdimensionale Modell: Für dieses fallen AIC und BIC am kleinsten aus. Es ist mit einer Wahrscheinlichkeit von (annähernd) 100 % ( $\omega_i$  und  $p_i$ ) das beste der drei Modelle (Tab. 3).

**Tab. 4:** Kennwerte für die Güte der drei berechneten Partial Credit-Modelle.

Modell	AIC	$\Delta AIC_i$	$\omega_i$	BIC	$\Delta BIC_i$	$p_i$
1D	13.622,99	174,91	0,00	13.916,01	107,66	0,00
2D	13.602,62	154,54	0,00	13.905,24	96,90	0,00
5D	13.448,08	0,00	1,00	13.808,34	0,00	1,00

Die fünf Dimensionen weisen Varianzen  $> 1$  (Tab. 4) und Interkorrelationen von 0,38 bis 0,77 auf. Die Regressionskonstanten zeigen an, dass alle Aufgaben im Mittel eher leicht sind, besonders aber diese der Teilkompetenzen Testen und Ändern von Modellen (Tab. 4).

**Tab. 5:** Varianzen und Regressionskonstanten (mit Standardabweichung) der fünf Dimensionen. Die Dimensionen sind nach den Teilkompetenzen Eigenschaften von Modellen (E), Alternative Modelle (A), Zweck (Z), Testen (T) und Ändern (Ä) von Modellen benannt.

	1: E	2: A	3: Z	4: T	5: Ä
<b>Varianz</b>	1,17	3,43	1,70	2,99	2,99
<b>Konstante</b>	0,60 (0,04)	0,56 (0,06)	0,42 (0,04)	1,83 (0,06)	1,32 (0,06)

Unter Verwendung des fünfdimensionalen Modells ergeben sich für die einzelnen Aufgaben wMNSQ-Werte im Bereich  $0,93 \leq wMNSQ \leq 1,13$ , für die Schwellenparameter der einzelnen Aufgaben Werte im Bereich  $0,89 \leq wMNSQ \leq 1,15$ . Es resultieren T-Werte von  $-0,8 \leq T \leq 1,3$  bzw.  $-1,8 \leq T \leq 1,4$ . Dies impliziert eine gute Passung der Aufgaben mit dem Modell (BOND & FOX, 2001).

## 5.2 Fragestellung 2: Prüfung inhaltsbezogener Teilkompetenzen

Sowohl der Modalwert als auch die Spannweite beträgt bei allen Aufgaben zwei. Hierbei wird Niveau I in vielen Fällen von weniger als 10 % der Schüler präferiert. Einen ersten Einblick in die Übereinstimmung der Schülerantworten bei den unterschiedlichen Aufgaben bietet Cramer's V. Für 89 % aller Vergleiche ist V signifikant und nimmt einen Wert im mittleren ( $0,3 < V < 0,5$ ) oder hohen Bereich an ( $0,5 < V$ ; nach FIELD, 2009). Somit korrelieren die Schülerantworten in 89 % der eingesetzten Aufgaben untereinander signifikant.

**Tab. 6:** Kennwerte für den Vergleich des eindimensionalen Modells mit dem dreidimensionalen Modell.

Modell	AIC	$\Delta AIC_i$	$\omega_i$	BIC	$\Delta BIC_i$	$p_i$
3D	13.624,38	176,30	0,00	13.941,41	133,07	0,00
5D	13.448,08	0,00	1,00	13.808,34	0,00	1,00

Um die Daten auf inhaltsbezogene Teilkompetenzen zu prüfen, wurde in ConQuest ein dreidimensionales Modell mit den drei Inhaltsbereichen als jeweils separate Dimension spezifiziert. Der Vergleich des fünfdimensionalen, auf die Teilkompetenzen bezogenen Modells, mit dem dreidimensionalen Modell zeigt, dass das dreidimensionale Modell keine bessere Erklärung der Daten liefert (Tab. 5). Ein Vergleich mit den absoluten Werten (AIC und BIC) des ein- und zweidimensionalen Modells (Tab. 3) zeigt darüber hinaus, dass das dreidimensionale Modell die Daten am schlechtesten beschreibt.

Schließlich wurden die drei Inhaltsbereiche als mögliche Prädiktoren der Aufgabenschwierigkeit in einem Regressionsmodell geprüft. In der Regressionsanalyse (Methode: Einschluss) ergeben sich signifikante Unterschiede zwischen den Inhaltsbereichen Humanbiologie und Botanik ( $\beta = 0,552$ ;  $p < 0,01$ ) nicht aber zwischen Humanbiologie und Zoologie ( $\beta = 0,359$ ;  $p > 0,05$ ). Aufgrund des negativen Regressionsgewichts des Prädiktors sind die humanbiologischen Aufgaben eher leicht. Insgesamt lassen sich etwa 24 % der Varianz der Aufgabenschwierigkeit durch die drei Inhaltsbereiche erklären (Tab. 6; Analyse 1). Da dieses Ergebnis gegen das Resultat der Modellprüfung mit ConQuest spricht, wonach keine inhaltsbezogenen Teilkompetenzen vorliegen, wurde eine zweite Regressionsanalyse mit den sechs Modellen (vgl. Tab. 2) als Prädiktoren durchgeführt. Wird hierbei das (humanbiologische) Mundmodell als Vergleichsstandard gewählt, ergeben sich signifikante Regressionsgewichte für drei Modelle (Samen, Palme, Libelle; jeweils  $p < 0,05$ ). Das Regressionsmodell erklärt knapp 30 % der Varianz der Aufgabenschwierigkeit (Tab. 6; Analyse 2). Mit dem zweiten humanbiologischen Modell als Vergleichsstandard (Fußmodell) ergeben sich keinerlei signifikante Regressionsgewichte (nicht tabellarisch dargestellt).

**Tab. 7:** Regressionsgewicht (B), Standardabweichung des Regressionsgewichts (SD B) und standardisiertes Regressionsgewicht ( $\beta$ ) beider Regressionsanalysen. \* bzw. \*\*: Wert ist auf dem Niveau  $< 0,05$  bzw.  $< 0,01$  signifikant.

Prädiktoren	B	SD B	$\beta$
<b>Analyse 1: 3 Inhaltsbereiche (<math>R^2 = 0,236</math>)</b>			
(Humanbiologie)	-0,123	0,055	---
Botanik	0,223	0,078	0,552**
Zoologie	0,145	0,078	0,359
<b>Analyse 2: 6 Modelle (<math>R^2 = 0,299</math>)</b>			
(Mund)	-0,155	0,080	---

Fuß	0,065	0,113	0,128
Samen	0,238	0,113	0,465*
Palme	0,274	0,113	0,536*
Libelle	0,251	0,113	0,492*
Vogel	0,104	0,113	0,204

## 6 Diskussion der Ergebnisse

Bezüglich der Dimensionalitätsprüfung mit ConQuest sprechen die Daten für die fünfdimensionale Lösung, das heißt für die Existenz von fünf konstruktbezogenen Teilkompetenzen im Sinne der Teilkompetenzen des Kompetenzmodells der Modellkompetenz. Das heißt, die Schülerantworten fallen in den fünf Teilkompetenzen potenziell verschieden aus. Dies wird zusätzlich durch die unterschiedlich ausfallenden Regressionskonstanten (Tab. 4) und die vorwiegend im mittleren Bereich liegenden Korrelationen zwischen den Teilkompetenzen gestützt. Die Kennwerte  $\omega_i$  und  $p_i$  für den Modellvergleich zeichnen mit einer Wahrscheinlichkeit von (annähernd) 100 % ein sehr deutliches Bild zugunsten des fünfdimensionalen Modells. Angesichts der absoluten Werte (AIC bzw. BIC) fällt die Berechnungsgrundlage von  $\omega_i$  und  $p_i$  (d. h.  $\Delta AIC_i$  bzw.  $\Delta BIC_i$ ) aber eher gering aus (Tab. 3). Um bezüglich der konstruktbezogenen Teilkompetenzen Sicherheit zu bekommen, sind daher bereits weitere FC-Aufgaben in einer größeren Stichprobe im Einsatz.

Fragestellung 2 bezieht sich auf die Überprüfung inhaltsbezogener Teilkompetenzen. Die Hypothese, nach der die unterschiedlichen Inhalte keinen wesentlichen Einfluss auf die Schülerantworten haben, kann bestätigt werden, wird Cramer's V als Maß der Korrelation herangezogen. Der Modellvergleich mit dem ein- und dem dreidimensionalen Modell zeigt darüber hinaus, dass die Berücksichtigung der drei Inhaltsbereiche keine verbesserte Erklärung der Datenstruktur liefert. Die Regressionsanalyse hat Unterschiede zwischen den humanbiologischen und botanischen, nicht aber zwischen den humanbiologischen und zoologischen Aufgaben ergeben. Die Analyse des Regressionsgewichts der einzelnen Modelle (d. h. Aufgabenstämme) weist auf Unterschiede zwischen dem Mundmodell und drei anderen Modellen (Samen, Palme, Libelle) hin. Da sich keine analogen Effekte des Fußmodells zeigen, sprechen diese Ergebnisse für einen Einfluss des Mundmodells auf die Aufgabenschwierigkeit und nicht für einen Einfluss des Inhalts Humanbiologie. Das heißt, die biologischen Disziplinen beeinflussen nicht systematisch die Schülerantworten. Aufgrund des

gewählten Facettendesigns mit stammäquivalenten Aufgaben zeigen diese Resultate aber, dass sich einzelne, in den Aufgabenstämmen eingesetzte Modelle über die Teilkompetenzen hinweg schwierigkeiterzeugend auswirken können.

Die Aussagekraft der Regressionsanalyse wird dadurch relativiert, dass die generellen Itemparameter (und nicht die Schwellenparameter) einbezogen und darüber hinaus deren Standardabweichungen nicht berücksichtigt werden konnten. Hierdurch beruht die Regressionsanalyse auf Werten, die in zweifacher Hinsicht lediglich eine Annäherung an die tatsächliche Datenstruktur darstellen. Die wMNSQ-Werte zeigen allerdings, dass die geschätzten Itemparameter eine gute Erklärung der tatsächlichen Datenstruktur liefern. Da für die Regressionsanalyse außerdem die Itemparameter des 5D-Modells gewählt wurden und in diesen der Einfluss der (konstruktbezogenen) Teilkompetenzen bereits modelliert ist (WU et al., 2007), kann das Regressionsgewicht der einzelnen Modelle überschätzt sein. Ungeachtet dieser methodischen Einwände sprechen die Ergebnisse der Regressionsanalyse dafür, dass einzelne Modelle schwierigkeiterzeugende Aufgabenmerkmale transportieren können. Hinweise hierauf wurden bereits andernorts berichtet (GROSSLIGHT et al., 1991, KRELL & KRÜGER, 2010, GRÜNKORN et al., im Druck). Da in der vorliegenden Studie Modelle eingesetzt wurden, die sich übereinstimmend auf einen Struktur-Funktions-Zusammenhang beziehen, ist dieses Ergebnis unerwartet bzw. deutet darauf hin, dass ‘tasks which may be similar to scientists may not appear that way to students’ (Clough & Driver, 1986, p. 475). Daher sollte der Einfluss des Aufgabenstamms auf die Schülerantworten bei den FC-Aufgaben zur Modellkompetenz weitergehend untersucht werden, denn: „Task contextuality is an issue of considerable importance for science assessment, perhaps more so in biology than in other science domains.” (NEHM & HA, 2011: 239).

## 7 Fazit & Ausblick

Es ist gezeigt worden, dass das fünfdimensionale Partial Credit-Modell am besten geeignet ist, die vorliegenden Daten zu erklären. Dieser Befund deckt sich mit der theoretisch erarbeiteten Struktur des Kompetenzmodells der Modellkompetenz (UPMEIER ZU BELZEN & KRÜGER, 2010). Da jedes Aufgabenformat spezifische kognitive Anforderungen transportiert (MARTINEZ, 1999), wird die Dimensionalität des Kompetenzmodells auch mit anderen methodischen Zugängen empirisch überprüft (Multiple Choice-Aufgaben: Terzer & Upmeier zu Belzen, 2010; offene Aufgaben: Grünkorn, Upmeier zu Belzen & Krüger, im Druck). Erst die gemeinsame und verbindende Analyse der Daten wird erge-

ben, in welcher Struktur sich Modellkompetenz abbilden lässt. Eine Studie, bei der die drei Aufgabenformate gemeinsam eingesetzt wurden, wird ihren direkten Vergleich und damit einen Einblick in die Abhängigkeit der Schülerantworten vom Aufgabenformat erlauben (psychometrische Aufgabenanalyse; MARTINEZ, 1999).

Nach der vorliegenden Studie stellt der fachliche Inhalt kein wesentliches schwierigkeiterzeugendes Merkmal der FC-Aufgaben dar. Es gibt allerdings einen Einfluss eines in den Aufgaben präsentierten Modells. Um diesen Einfluss besser zu verstehen, wird ein qualitativer Ansatz verfolgt: Orientiert an der Repertory Grid-Methode ordnen und bewerten Schüler unterschiedliche biologische Modelle. Auf diesem Weg wird eine schülerbasierte Modell-Klassifikation erarbeitet. In weiteren Untersuchungen wird darauf aufbauend analysiert, inwieweit sich gleich klassifizierte Modelle schwierigkeiterzeugend auswirken und als merkmalsbezogene Teilkompetenz berücksichtigt werden sollten.

## Zitierte Literatur

- BLUM, A., H. GOLDSTEIN & F. GUÉRIN-PACE (2001): International Adult Literacy Survey (IALS): an analysis of international comparisons of adult literacy. *Assessment in Education* 8, 225-246.
- BOND, T.G. & C.M. FOX (2001): *Applying the Rasch Model: Fundamental Measurement in the Human Science*. Lawrence Erlbaum, Mahwah.
- BORTZ, J., G.A. LIENERT & K. BOEHNKE (2008): *Verteilungsfreie Methoden in der Biostatistik*. 3. korrigierte Auflage. Springer, Heidelberg.
- BURNHAM, K.P. & D.R. ANDERSON (2004): Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* 33, 261-304.
- CHAN, W. (2003): Analyzing Ipsative Data in Psychological Research. *Behaviormetrika* 30, 99-121.
- CLOUGH, E. E. & DRIVER, R. (1986): A Study of Consistency in the Use of Students' Conceptual Frameworks Across Different Task Contexts. *Science Education* 70, 473-496.
- COOMBS, C.H. (1950): Psychological Scaling without a Unit of Measurement. *Psychological Review* 57, 145-158.
- FIELD, A. (2009): *Discovering Statistics Using SPSS*. 3rd Edition. Sage, Los Angeles [u.a.].
- GILBERT, J.K. (2006): On the Nature of "Context" in Chemical Education. *International Journal of Science and Mathematics Education* 28, 957-976.
- GOLDSTEIN, H. (2004): International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education* 11, 319-330.
- 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, 799-822.
- GRÜNKORN, J., A. UPMEIER ZU BELZEN & D. KRÜGER (im Druck): Referierter Tagungsband zur 8th Conference of European Researchers in Didactics of Biology, Braga (Portugal), 13.07.-17.Juli 2010. 12 Seiten.
- HICKS, L.E. (1970): Some Properties of Ipsative, Normative, and Forced-Choice Normative Measures. *Psychological Bulletin* 74, 167-184.
- KAUERTZ, A. (2008): *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*. Logos, Berlin.

- KAUERTZ, A., H.E. FISCHER, J. MAYER, E. SUMFLETH & M. WALPULSKI (2010): Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften* 16, 135-153.
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland; Hrsg.) (1989/2004): Einheitliche Prüfungsanforderungen in der Abiturprüfung Biologie. (Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 05.02.2004). [[http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/1989/1989\\_12\\_01-EPA-Biologie.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Biologie.pdf); 12.05.2011]
- KMK (Hrsg.) (2005): Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Wolters Kluwer, München & Neuwied.
- KRELL, M. & D. KRÜGER (2010): Diagnose von Modellkompetenz. Deduktive Konstruktion und Selektion von geschlossenen Items. *Erkenntnisweg Biologiedidaktik* 9, 23-37.
- KRELL, M., A. CZESKLEBA, A. UPMEIER ZU BELZEN & D. KRÜGER (eingereicht): Validierung von Forced Choice-Aufgaben zur Modellkompetenz durch Lautes Denken – Ein Beitrag zur Integration quantitativer und qualitativer Methoden. 21 Seiten.
- MAHR, B. (2008): Ein Modell des Modellseins. Ein Beitrag zur Aufklärung des Modellbegriffs. In: Dirks, U. & E. Knobloch [Hrsg.]: *Modelle*. Peter Lang, 187-218.
- MARTINEZ, M.E. (1999): Cognition and the Question of Test Item Format. *Educational Psychologist* 34, 207-218.
- MASTERS, G.N. (1982): A Rasch Model for Partial Credit Scoring. *Psychometrika* 47, 149-174.
- NEHM, R.H. & M. HA (2011): Item Feature Effects in Evolution Assessment. *Journal of Research in Science Teaching* 48, 237-256.
- NEUMANN, K, A. KAUERTZ, A. LAU, H. NOTARP & F.E. FISCHER (2007): Die Modellierung physikalischer Kompetenz und ihrer Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften* 13, 101-121.
- PRENZEL, M., P. HÄUBLER, J. ROST & M. SENKBEIL (2002): Der PISA-Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten vorhersagen? *Unterrichtswissenschaft* 30, 120-135.
- ROST, J., O. WALTER, C.H. CARSTENSEN, M. SENKBEIL, M. PRENZEL (2005): Der nationale Naturwissenschaftstest PISA 2003. *MNU* 58, 196-204.
- SCHECKER, H. & I. PARCHMANN (2006): Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften* 12, 45-66.
- SCHMIEMANN, P. (2010): Modellierung von Schülerkompetenzen im Bereich des biologischen Fachwissens. Logos, Berlin.
- SENKBEIL, M., J. ROST, C.H. CARSTENSEN & O. WALTER (2005): Der nationale Naturwissenschaftstest PISA 2003. Entwicklung und empirische Überprüfung eines zweidimensionalen Facettendesigns. *Empirische Pädagogik* 19, 166-189.
- STARK, S., O.S. CHERNYSHENKO & F. DRASGOW (2005): An IRT Approach to Constructing and Scoring Pairwise Preference Items Involving Stimuli on Different Dimensions: The Multi-Unidimensional Pairwise-Preference Model. *Applied Psychological Measurement* 29, 184–203.
- TERZER, E. & UPMEIER ZU BELZEN, A. (2010). Model Competence in Biology Education: Operationalization and Validation of a Theoretical Model of Model Competence using Multiple-Choice Items. In A. Yarden, D. J. Boerwinkel, M. Ekborg, M. Reiss, V. Zogza, G. Carvalho,...(Hrsg.), *Eight Conference of European Resaerchers in Didactics of Biology (ERIDOB)* (S. 160). Braga: Universidade do Minho.
- TRAUB, R.E. & K. MACRURY (1990): Antwort-Auswahl- vs Freie-Antwort-Aufgaben bei Lernerfolgstests. In: K. Ingenkamp & R.S. Jäger [Hrsg.]: *Test und Trends* 8. Jahrbuch der Pädagogischen Diagnostik. Beltz, 128-159.
- UPMEIER ZU BELZEN, A. & D. KRÜGER (2010): Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften* 16, 41-57.
- VON DAVIER, A. A., CARSTENSEN, C. H. & VON DAVIER, M. (2008): Linking Competencies in Horizontal, Vertical, and Longitudinal Settings and Measuring Growth. In: Hartig, J., Klieme, E. & Leutner, D. (Hrsg.): *Assessment of Competencies in Educational Contexts*. Hogrefe & Huber, 121-149.
- WU, M. L., ADAMS, R. J., & WILSON, M. R. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell, Vic: ACER Press.
- YEN, W.M. & FITZPATRICK, A.R. (2006): Item Response Theory. In: Brennan, R.L. (Hrsg.): *Educational Measurement*. 4. Ed. Praeger, 111-153.