

Validierung von Forced Choice-Aufgaben durch lautes Denken

Moritz Krell, Anja Czeskleba & Dirk Krüger

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

Freie Universität Berlin, Didaktik der Biologie

Schwendenerstraße 1, 14195 Berlin

Zusammenfassung

Mit der outcome-Orientierung des deutschen Bildungswesens besteht eine zentrale Aufgabe empirischer Bildungsforschung in der Definition messbarer Kriterien dafür, was Schülerinnen und Schüler am Ende ihrer Schulzeit wissen und können sollen. Eine gängige Methode der Kompetenzdiagnostik beschreibt Kompetenzniveaus auf der Grundlage von Ergebnissen von Testaufgaben und bedient sich somit vor allem quantitativer Methoden. Der vorliegende Artikel unterstreicht die Notwendigkeit einer Verschränkung quantitativer und qualitativer Methoden und präsentiert anhand der inhaltlichen Validierung von Forced Choice-Aufgaben, die zur Beschreibung des Modellverständnisses von Schülerinnen und Schülern eingesetzt werden sollen, ein Beispiel der konkreten Umsetzung.

Abstract

It is a central challenge of empirical educational research to define measurable criteria for what students should know and what they should be able to do after finishing school. The classical method for the diagnosis of competences develops levels of competence based on the results of assessment and is therefore dominated by quantitative methods. The present article underlines the need for a collaboration of both quantitative and qualitative methods. An example of a mixed method-approach is given along with the presentation of the content validation of forced choice-items are developed as diagnostic tools for model competence.

1 Einleitung

Die Hinwendung zur outcome-Orientierung im deutschen Bildungswesen hat Fragen der Kompetenzdiagnostik und darauf aufbauende Fördermaßnahmen in den Fokus der Bildungsforschung gerückt (AUFSCHNAITER & ROGGE, 2010). Eine zentrale Aufgabe der empirischen Bildungsforschung besteht seither darin, messbare Kriterien für das zu definieren, was in den jeweiligen Bildungsinstitutionen vermittelt werden soll (HARTIG et al., 2008). In diesem Zusammenhang hat sich eine gängige Methode der Kompetenzmodellierung etabliert. Diese besteht im Kern darin, empirisch auf der Basis von Testaufgaben und mit Hilfe probabilistischer Statistik inhaltliche Kompetenzstufen herauszuarbeiten („formale Stufenbeschreibung“, FRANKE, 2005; RUPP et al., 2006). Demgegenüber arbeitet vorliegende Studie mit einem theoretisch erarbeiteten Kompetenzmodell (UPMEIER ZU BELZEN & KRÜGER, 2010). Wichtig bei beiden Vorgehen ist nicht nur, die kriteriale Validität der Instrumente zu belegen (KUNTER & KLUSMANN, 2010). Auch die Prüfung der inhaltlichen Validität der Diagnoseaufgaben als eine genuin qualitative Frage sollte nicht vernachlässigt werden (KELLE & REITH, 2008).

Die Vorzüge einer integrativen Umsetzung quantitativer und qualitativer Methoden werden von vielen Autoren zwar betont, dies hat aber „wenig beigetragen [...] zur Überwindung des grundsätzlichen Schismas zwischen der qualitativen und quantitativen Methodentraktion“ (KELLE, 2008, S. 48). Die vorliegende Untersuchung führt anhand eines Beispiels die praktische Umsetzung eines *mixed method*-Designs aus (FOSCHT et al., 2007; TASHAKKORI & TEDDLIE, 2003) und leistet auf diese Weise einen exemplarischen Beitrag zur Integration quantitativer und qualitativer Methoden in der wissenschaftlichen Praxis: Auf der theoretischen Grundlage eines Kompetenzmodells für Modellkompetenz im Biologieunterricht (UPMEIER ZU BELZEN & KRÜGER, 2010) wurden deduktiv Items konstruiert und anhand quantitativer Analysen selektiert (KRELL & KRÜGER, 2010). Diese Items bilden die Grundlage von Forced Choice-Aufgaben zur Diagnose des Modellverständnisses von Schülerinnen und Schülern (KRELL & KRÜGER, 2011). Das im vorliegenden Artikel beschriebene Ziel ist die inhaltliche Validierung der Forced Choice-Aufgaben mit Hilfe des Lauten Denkens (ERICSSON & SIMON, 1980). Damit wird die praktische Umsetzung eines (parallelen) *mixed method*-Designs ausgeführt (FOSCHT et al., 2007): Quantitative und qualitative Methoden werden parallel eingesetzt, um eine inhaltliche Validierung der Aufgaben aus Sicht der Probanden zu erreichen.

2 Theoretischer Rahmen

2.1 Modellkompetenz im Biologieunterricht

Um Schülerinnen und Schüler gezielt fördern zu können, sind Kompetenzmodelle hilfreich, die auf die inhaltliche Struktur und die mögliche Entwicklung von (kognitiven) Voraussetzungen für das Lösen von Problemen eingehen (HARTIG et al., 2008; SCHECKER & PARCHMANN, 2006). In diesem Zusammenhang haben Upmeier zu Belzen und Krüger (2010) auf der Grundlage diverser Studien (z. B. CRAWFORD & CULLIN, 2004; GROSSLIGHT et al., 1991; JUSTI & GILBERT, 2003) und unter Einbezug der theoretischen Erörterungen MAHRS (2008) ein Kompetenzmodell der Modellkompetenz entwickelt. In diesem wird das Verständnis von Schülerinnen und Schülern über Modelle in den Dimensionen Kenntnisse über Modelle und Modellbildung beschrieben. Die erste Dimension beschreibt abstrakte Konzepte über Modelle und ist inhaltlich in die Teilkompetenzen *Eigenschaften von Modellen* und *Alternative Modelle* differenziert. Erstere bezieht sich auf das Verhältnis zwischen Modell und Original. Die zweite Perspektive betrachtet mögliche Gründe für unterschiedliche Modelle eines Originals. Die Dimension Modellbildung fokussiert auf den Prozess der Modellierung und ist in drei Teilkompetenzen differenziert, die jeweils unterschiedliche Perspektiven auf den *Zweck*, das *Testen* und das *Ändern* von Modellen beschreiben. Jede Teilkompetenz kann in drei unterschiedlichen Komplexitätsniveaus ausgeprägt sein, die sich, vereinfacht ausgedrückt, darin unterscheiden, dass auf Niveau I und II das Modell aus der Herstellungsperspektive als Nachbildung des Originals zu Veranschaulichungs- oder Erklärungszwecken betrachtet wird, während es auf Niveau III aus der Anwendungsperspektive als hypothetisch angesehen wird, Untersuchungen notwendig macht und damit neue Erkenntnisse ermöglicht (für eine differenziertere Beschreibung siehe UPMEIER ZU BELZEN & KRÜGER, 2010). Das Kompetenzmodell der Modellkompetenz beschreibt Facetten von Modellverständnis und strukturiert somit die Voraussetzungen für einen kompetenten Umgang mit Modellen. Diese Struktur muss zunächst empirisch validiert werden, um auf der Grundlage eines validierten Kompetenzmodells Diagnoseaufgaben entwickeln zu können, die das Verständnis von Schülerinnen und Schülern über Modelle und das Modellieren erfassen können und somit eine individuelle Förderung unterstützen. KRELL & KRÜGER (2010) haben in einem ersten Schritt deduktiv Items für jede Zelle des Kompetenzmodells entwickelt, um diese in Forced Choice-Aufgaben zur Beschreibung des Modellverständnisses von Lernenden implementieren zu können (KRELL & KRÜGER, 2011).

2.2 Forced Choice-Aufgaben

Der Begriff Forced Choice (FC) beschreibt geschlossene Aufgaben, bei denen sich die Probanden zwischen mehreren zur Verfügung stehenden Antwortalternativen entscheiden müssen. Ein typisches Beispiel für FC-Aufgaben sind (Dominanz-)Paarvergleiche, bei denen mehrmals zwei Antwortalternativen gegenüber gestellt werden und Probanden sich für eine der beiden Alternativen entscheiden müssen (Abb. 1). Die für einen vollständigen Vergleich aller Antwortalternativen (n) erforderlichen Paarvergleichsurteile nehmen mit wachsendem n schnell zu, weshalb Paarvergleiche einen hohen *cognitive load* transportieren können. Brown & Bartram (2009) empfehlen daher, nicht mehr als drei Antwortalternativen in eine FC-Aufgabe zu integrieren.

Der **Neandertaler** ist ein früher Verwandter des heutigen Menschen. Es wird geschätzt, dass der Neandertaler vor etwa 25.000 Jahren ausgestorben ist.

Aus Knochenfunden und wissenschaftlichen Untersuchungen haben ForscherInnen ein **Modell des Neandertalers** entwickelt. Abbildung 1 (links) zeigt den Schädel eines Neandertalers, der bei Ausgrabungen in Frankreich gefunden wurde. Abbildung 2 (rechts) zeigt ein Modell eines Neandertalerkopfes.



Abb. 1: Knochenfund



Abb. 2: Modell

Was denkst Du über das Modell des Neandertalerkopfes?

Lies die folgenden Aussagen. Markiere in jeder Zeile diejenige, die deiner Meinung nach stärker auf das Modell des Neandertalerkopfes zutrifft. Mache in jeder Zeile 1 Kreuz (X)!

Welche Aussage trifft stärker zu? Mache in jeder Zeile nur ein X.	
Das Modell des Neandertalerkopfes ...	
... macht wesentliche Merkmale des Neandertalerkopfes sachgemäß deutlich. <input type="checkbox"/>	... bildet ab, welche Annahmen man über den Neandertalerkopf hat. <input type="checkbox"/>
... sieht so aus wie der Neandertalerkopf. <input type="checkbox"/>	... macht wesentliche Merkmale des Neandertalerkopfes sachgemäß deutlich. <input type="checkbox"/>
... bildet ab, welche Annahmen man über den Neandertalerkopf hat. <input type="checkbox"/>	... sieht so aus wie der Neandertalerkopf. <input type="checkbox"/>

Abb. 1: Paarvergleich für die Teilkompetenz Eigenschaften von Modellen. In der linken Spalte repräsentieren die Formulierungen Niveau II, Niveau I und Niveau III (von oben nach unten).

individuelle Normen und keine objektiven Skalenpunkte die Grundlage für die Bearbeitung von FC-Aufgaben sind, entstehen relative Urteile („ipsative Daten“; HICKS, 1970). Daher ist ein interindividueller Vergleich von Ergebnissen,

Einen theoretische Ansatz, welche kognitiven Entscheidungsprozesse für das Bearbeiten einer FC-Aufgabe notwendig sind, hat COOMBS (1950) vorgelegt. Demnach fällt die Entscheidung zwischen zwei Antwortalternativen stets auf diejenige, die näher an der eigenen Vorstellung („individual ideal or norm“; COOMBS, 1950, S. 146) liegt. Bei einer Serie von Paarvergleichsurteilen wird demnach jedes Mal erneut dahingehend geurteilt, welche der beiden Alternativen stärker der eigenen Vorstellung entspricht. Aus diesen Überlegungen lässt sich ein wichtiges Charakteristikum der durch FC-Aufgaben generierten Daten ableiten: Da

die mit Hilfe von FC-Aufgaben gewonnen werden, problematisch (BÖCKENHOLT, 2004; CHAN, 2003). Nichtsdestotrotz beschäftigen sich Studien mit statistischen Möglichkeiten, die Ergebnisse von FC-Aufgaben auch interindividuell vergleichen zu können (BROWN & MAYDEU-OLIVERS, 2011; MCCLOY et al., 2005).

FC-Aufgaben besitzen spezifische Charakteristika, die sie positiv auszeichnen: Zunächst bedeutet der hohe *cognitive load* bei der Bearbeitung von FC-Aufgaben (BROWN & BARTRAM, 2009) theoretisch eine intensive Auseinandersetzung mit den jeweiligen Antwortalternativen. Außerdem umgeht der direkte Vergleich zwischen den Antwortalternativen Probleme, die aus der individuellen Interpretation vorgegebener Skalenpunkte resultieren können (BÖCKENHOLT, 2004; CLANCY & GARSEN, 1970). Auch das Problem der gleichen Bewertung mehrerer Antwortalternativen wird vermieden (BÖCKENHOLT, 2004). Schließlich lässt sich über eine Analyse zirkulärer Triaden (bei Paarvergleichen) auf die Konsistenz des Antwortverhaltens schließen (BÖCKENHOLT, 2004; BORTZ et al., 2008). FC-Aufgaben wurden daher bereits für die Beschreibung des Modellverständnisses von Schülerinnen und Schülern (CHITTLEBOROUGH et al., 2005) sowie für die Modellierung naturwissenschaftlicher Kompetenz von Grundschülerinnen und Grundschülern verwendet (KLEICKMANN et al., 2010).

2.3 FC-Aufgaben für die Beschreibung von Modellverständnis

Das Ergebnis einer Bewertung unterschiedlicher Ausprägungen eines eindimensionalen Konstrukts mit Hilfe von FC-Aufgaben kann als individuelle Positionierung auf eben diesem Konstrukt interpretiert werden (MCCLOY et al., 2005). Die Frage der Dimensionalität des Kompetenzmodells der Modellkompetenz ist Gegenstand laufender Studien und noch nicht abschließend empirisch geklärt (KRELL & KRÜGER, 2011; TERZER & UPMEIER ZU BELZEN, 2010). Unabhängig davon produzieren die FC-Aufgaben nach MCCLOY et al. (2005) für jede Teilkompetenz eine eindeutige Rangfolge der drei Aussagen, die somit als intrapersonale Präferenzrangfolge in Bezug auf diese drei Aussagen interpretiert werden kann. Schülerinnen und Schüler, die bei einer Aufgabe beispielsweise zwei Mal die Aussage von Niveau II und ein Mal die Aussage von Niveau I ausgewählt haben, zeigen demnach bei dieser Aufgabe ein Modellverständnis, das zwischen Niveau I und II – mit einer Präferenz für Niveau II – lokalisiert ist.

2.4 Modelle in der Schule und in der Wissenschaft

In der Literatur existieren unterschiedliche Vorschläge zur Systematisierung von Modellen (z. B. GILBERT et al., 2000; HARRISON & TREAGUST, 2000). In der biologiedidaktischen Literatur wird u. a. die Unterscheidung zwischen Schul- und Wissenschaftsmodellen diskutiert (KATTMANN, 2006). *Schulmodelle* repräsentieren Erkenntnisse, die in der Wissenschaft bereits anerkannt sind. Im Vergleich zu wissenschaftlichen Modellen stellen Schulmodelle biologische Phänomene häufig vereinfacht dar. Modelle werden in Schule häufig zur Darstellung oder Erklärung eingesetzt (VAN DRIEL & VERLOOP, 2002). Schulmodelle können aus Sicht von Lernenden aber unbekannte Tatsachen repräsentieren und zur Erkenntnisgewinnung eingesetzt werden. *Wissenschaftsmodelle* beziehen sich auf Zusammenhänge, die Gegenstand aktueller Forschung sind und dienen in diesem Rahmen als Mittel der Erkenntnisgewinnung (GILBERT et al., 2000).

Trier und UPMEIER ZU BELZEN (2009) haben gezeigt, dass Schülerinnen und Schüler zwischen der Funktion von Modellen in der Schule und in der Wissenschaft differenzieren. Folglich kann das Modellverständnis von Lernenden davon beeinflusst sein, ob sie ein gegebenes Modell eher mit wissenschaftlichen oder schulischen Kontexten verbinden. Für die vorliegende Studie wurden von den Autoren und der Autorin sechs Modelle ausgewählt (Abb. 1, Abb. 2). Es wird davon ausgegangen, dass sie von Schülerinnen und Schülern eher mit Schule (Abb. 2: (1), (2), (3)) beziehungsweise eher mit Wissenschaft (Abb. 2: (4), (5); Abb. 1) assoziiert werden.

3 Fragestellungen

Das primäre Ziel vorliegender Untersuchung ist die inhaltliche Validierung der FC-Aufgaben zur Beschreibung des Modellverständnisses von Schülerinnen und Schülern. Die für die Untersuchung leitende Fragestellung ist daher:

- (1) Inwiefern entspricht die Interpretation der Formulierungen in den FC-Aufgaben durch die Probanden während der Aufgabenbearbeitung inhaltlich dem beabsichtigten Niveau im Kompetenzmodell?

FC-Aufgaben verlangen von den Probanden spezifische Entscheidungsprozesse und resultieren theoretisch in individuellen Präferenzrangfolgen der präsentierten Antwortalternativen (vgl. 2.2, 2.3). Die Äußerungen der Schülerinnen und Schüler können dahingehend analysiert werden, ob diese theoretischen Beschreibungen zutreffen. Diesbezüglich wird folgende Fragestellung diskutiert:

- (2) Inwiefern erlauben die FC-Aufgaben eine inhaltlich valide Feststellung der ersten und zweiten Präferenz der Probanden bezüglich der drei Niveaus einer Teilkompetenz?

Schließlich kann durch den qualitativen Ansatz überprüft werden, ob das durch die Schülerinnen und Schüler geäußerte Modellverständnis in dem Kompetenzmodell der Modellkompetenz beschrieben ist. Eine weitere Fragestellung ist somit:

- (3) Entsprechen die Äußerungen der Probanden während der Aufgabenbearbeitung inhaltlich den Facetten des Kompetenzmodells der Modellkompetenz?

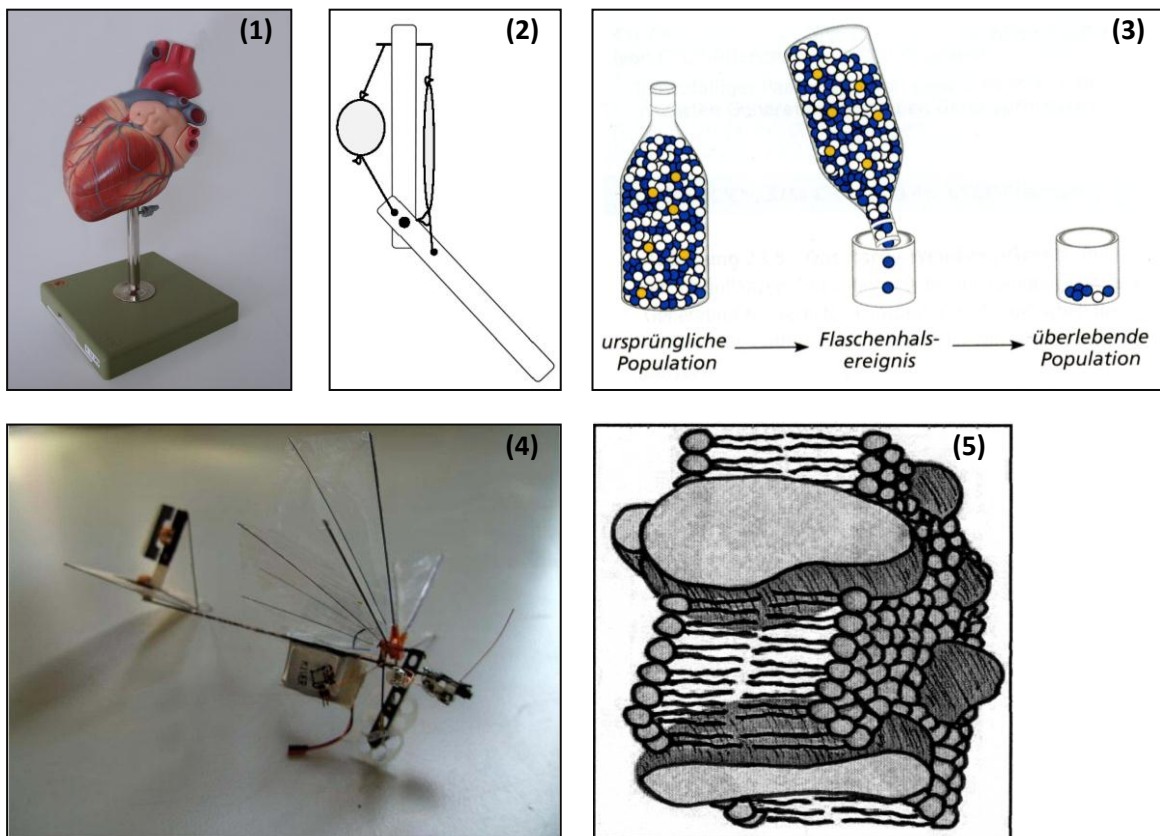


Abb. 2: Für die Studie ausgewählte Modelle. (1): Strukturmodell des menschlichen Herzens; (2): Funktionsmodell des menschlichen Arms; (3): Flaschenhalsmodell zur Modellierung von zufälligen Populationsveränderungen (Campbell und Reece, 2011, S. 306); (4): Technisches Modell einer Libelle (Eisma, 2012, S. 16); (5): Strukturmodell einer Biomembran.

4 Methodisches Vorgehen

Für die Beantwortung der Fragestellungen wurde nach der Methode des Lauten Denkens vorgegangen (ERICSSON & SIMON, 1980). Schülerinnen und Schüler

wurden dazu aufgefordert, alles laut auszusprechen, was ihnen beim Bearbeiten der FC-Aufgaben durch den Kopf geht. Gegebenenfalls wurde während der Bearbeitung erneut an die Aufgabe der Verbalisierung der Gedanken erinnert oder es wurde darum gebeten, die Gedankengänge, die zu bestimmten Entscheidungen geführt haben, nachträglich darzustellen (SCHOLL, 2009). Alle Probanden bearbeiteten zwei FC-Aufgaben zu jeder Teilkompetenz des Kompetenzmodells, wobei die sechs Modelle in der Weise verteilt wurden, dass für jede Teilkompetenz ein eher wissenschaftliches und ein eher schulisches Modell bearbeitet werden musste. Einige Modelle wurden dabei für zwei Aufgaben verwendet (*Eigenschaften von Modellen* und *Alternative Modelle: Herz- und Zellmembranmodell*; *Zweck und Testen von Modellen: Arm- und Libellenmodell*; *Ändern von Modellen: Umweltkatastrophen- und Neandertalermmodell*). Die Struktur der FC-Aufgaben ist Abbildung 1 zu entnehmen: In jeder Aufgabe wird mit einem kurzen Text der fachliche Hintergrund beschrieben und das Modell sowie ggf. das dazugehörige Original abgebildet. Anschließend folgt die eigentliche FC-Aufgabe, in der auf das jeweilige Modell Bezug genommen wird.

Das Sampling wurde nach einem „qualitativen Stichprobenplan“ (KELLE & KLUGE, 2010) gestaltet. Das heißt, es wurde vorab definiert, wie viele Probanden aus welcher Jahrgangsstufe befragt werden sollen. Nach dem umgesetzten Stichprobenplan wurden insgesamt elf Schülerinnen und Schüler der Klassenstufen 7 bis 13 eines Berliner Gymnasiums befragt (Klasse 7 und 8: je 3 Probanden, Kl. 9: 2, Kl. 10, 11 und 12: je 1). Es wurde bewusst ein Schwerpunkt auf die unteren Klassenstufen gelegt, da bei diesen größere Verständnisschwierigkeiten erwartet wurden. Die Verbalisierungen wurden nach GROPEN-GIEBER (2005) transkribiert, redigiert sowie deduktiv (MAYRING, 2000) auf der Grundlage eines von GRÜNKORN et al. (2011) erarbeiteten Codierschemas für das Kompetenzmodell der Modellkompetenz codiert.

Für die Beantwortung der Fragestellungen 1 und 2 wurden die redigierten Äußerungen inhaltlich interpretiert und den Niveaus des Kompetenzmodells zugeordnet. Für Fälle, bei denen eine Zuordnung nicht möglich war, wurden abduktiv⁷ neue Kategorien entwickelt, um einen Hinweis auf eventuelle Lücken im

⁷ Kelle & Kluge (2010) sprechen dezidiert von abduktiver, nicht von induktiver Codierung, und beschreiben das „induktivistische Selbstmissverständnis“ (S. 16): Demnach ist ein induktives Vorgehen für qualitative Forschung schlichtweg unmöglich, da von Forscherinnen und Forschern stets theoretisches Vorwissen eingebracht wird. Qualitatives Vorgehen zeichnet sich vielmehr durch einen „Zangengriff“ (S. 23) von Empirie und Theorie aus. Der Begriff der Abduktion geht hierbei auf Peirce

Kompetenzmodell zu erlangen (Fragestellung 3). Das intendierte und tatsächliche Verständnis der Formulierungen wurde in Kreuztabellen vergleichend gegenübergestellt, um auf diesem Wege typische Interpretationsprobleme aufdecken zu können. Kreuztabellen bieten einen Überblick über alle potentiellen und tatsächlichen Kombinationsmöglichkeiten und sind daher ein guter Ausgangspunkt, empirische Muster inhaltlich zu deuten und zu Typen zu verdichten (KELLE & KLUGE, 2010).

5 Darstellung der Ergebnisse

Im Folgenden werden die Ergebnisse nicht nach den in den Fragebögen eingesetzten Modellen separiert dargestellt, da sich keine systematischen modellabhängigen Unterschiede bezüglich der Interpretation der Formulierungen (Fragestellung 1), der Entscheidungsprozesse während der Aufgabenbearbeitung (Fragestellung 2) und zusätzlicher Perspektiven auf Modelle und das Modellieren (Fragestellung 3) gezeigt haben.

5.1 Validierung der FC-Aufgaben (Fragestellung 1)

Für die Beantwortung der Fragestellung 1 werden Kreuztabellen präsentiert (Tab. 1) und die idealtypischen Verständnisprobleme beschrieben. Aus dem Sampling der Untersuchung ergeben sich theoretisch 22 Äußerungen pro Niveau. Unterschiedliche Zeilensummen treten auf (Tab. 1), da nicht alle Äußerungen sicher interpretiert und zugeordnet werden konnten. Die Zahlen in den Diagonalen zeigen die Fälle, in denen die Aufgaben in der beabsichtigten Weise verstanden worden sind (grau hinterlegt). Im Folgenden werden einige Interpretationsschwierigkeiten exemplarisch beschrieben.

In der Teilkompetenz *Eigenschaften von Modellen* wurde ein Mal Niveau I und sieben Mal Niveau III nicht wie beabsichtigt interpretiert. Zum Beispiel entscheidet sich Proband 02G09 beim Vergleich zwischen Niveau III und Niveau I für Letzteres, interpretiert die Niveau I-Aussage *Das Modell des Herzens sieht so aus wie das Herz*. (vgl. Abb. 1) im Sinne von Niveau II: „Ich denke, es [...] sieht so aus wie das Herz. Zwar ist es nicht genau perfekt, aber es ist so ähnlich und das Modell ist der Durchschnitt sozusagen.“ Die sieben nicht intendierten Interpretationen von Niveau III als Niveau II gehen darauf zurück, dass die Probanden den hypothetischen Charakter der Niveau III-Aussage (*Das Modell vom Herzen bildet ab, welche Annahmen man über das echte Herz hat.*; vgl.

(z. B. 1967) zurück und beschreibt neben der Induktion und Deduktion einen dritten Weg des Schlussfolgerns.

Tab. 2) nicht so wie beabsichtigt verstanden haben, z. B.: „Annahmen, also was für Grundvoraussetzungen das Herz haben muss. [...] Und was das so hat und so.“ (13G07). Es wird deutlich, dass der hypothetische Charakter des Modells, der mit dem Begriff Annahmen intendiert wurde, nicht hinreichend verstanden wird. Auch in der Teilkompetenz *Alternative Modelle* legt z. B. Schüler 10G08 das Erforschen unterschiedlicher Vermutungen als schulisches Lernen aus und deutet die Niveau III-Aussage (*Es gibt verschiedene Modelle der Biomembran, weil mit ihnen unterschiedliche Vermutungen über die Biomembran erforscht werden sollen.*; vgl. Tab. 2) im Sinne von Niveau II: „Ich denke, dass verschiedene Vermutungen über die Biomembran erforscht werden sollen. Also, dass verschiedene Altersgruppen angesprochen werden.“ Schließlich treten solche Uminterpretationen von Niveau III in Niveau II auch bei den Teilkompetenzen *Zweck* und *Testen* von Modellen auf. Für die Teilkompetenz *Zweck* von Modellen wird die Aussage (*Das Libellenmodell hat den Zweck, neue Vermutungen über die Libelle zu erforschen.*; vgl. Tab. 2) von Schüler 10G08 wie folgt gedeutet: „Man soll es wahrscheinlich den Schülern näher bringen. Also die Libelle, wie sie funktioniert, wie die fliegt und so weiter.“ Und in der Teilkompetenz *Testen* von Modellen versteht Schüler 03G11 die Niveau III-Aussage (*Das Armmodell wird getestet, indem man testet, ob neue Erwartungen über den Arm zutreffen.*; vgl. Tab. 2) im Sinne von Niveau II: „Also der Arm, der muss sich so und so weit strecken können. [...] Und das sind dann sozusagen diese neuen, okay, jetzt nicht so neu, halt diese Erwartungen an dieses Modell.“

Schließlich erwiesen sich in einzelnen Fällen konkrete Begriffe als problematisch. In der Teilkompetenz *Zweck* von Modellen klingt die Formulierung von Niveau II sehr dinglich (vgl. Tab. 2) und hat sich damit zu weit von der theoretischen Bedeutung entfernt, die allgemein auf das Zusammenspiel verschiedener Variablen abzielt. Insbesondere der Begriff „Verknüpfungen“ hat sich hierbei als problematisch erwiesen. In der Teilkompetenz *Testen* von Modellen wurde in einigen Fällen der Begriff „Beschädigungen“ auf das Original und nicht auf das Modell bezogen.

Auch in der Teilkompetenz *Ändern* von Modellen warf der Begriff „Beschädigungen“ Probleme auf. Dieser wurde oft als Fehler, also als mangelnde Passung zwischen Original und Modell, ausgelegt.

Tab. 1: Intendierte Bedeutung und erfolgte Interpretation der Formulierungen. I, II, III: Niveaus im Kompetenzmodell. Für die Umdeutungen (U), in denen also Intention und Interpretation nicht übereinstimmen, ist die Anzahl (A) angegeben, die bei dem Vergleich in dem Paar geäußert wurde, das die erste Präferenz nicht enthält (A/U; vgl. 5.3).

Kenntnisse über Modelle										
		Eigenschaften von Modellen			Alternative Modelle					
		Interpretation								
		I	II	III	I	II	III			
Intention	I	10	1/1	-	3	1/2	2/2			
	II	-	21	-	-	16	1/1			
	III	-	6/7	9	-	0/2	14			

Modellbildung										
		Zweck von Modellen			Testen von Modellen			Ändern von Modellen		
		Interpretation								
		I	II	III	I	II	III	I	II	III
Intention	I	18	0/1	-	8	1/1	2/2	11	0/7	-
	II	1/1	15	-	-	19	-	-	13	1/3
	III	0/1	3/6	10	-	1/5	12	-	1/2	15

Die Formulierungen von Niveau II und III in der Teilkompetenz *Ändern* von Modellen wurden missverstanden, weil sie inhaltlich nicht gut voneinander getrennt werden konnten. Einige Schülerinnen und Schüler besaßen eine elaborierte Vorstellung von dem wissenschaftlichen Forschungsprozess und deuteten entweder die Formulierung von Niveau III oder Niveau II stärker in diesem Sinne. So auch Schülerin 02G09: „Es treffen beide Aussagen zu, das ist klar. Aber die zweite [Niveau II] trifft mehr zu. Weil man probiert ja dieses Modell aus, um auf ein Ergebnis zu kommen und mit diesen Ergebnissen arbeitet man dann ja weiter.“ Die Schülerin entscheidet sich in dieser Situation für Niveau II, argumentiert aber auf Niveau III. Diese Entscheidung wird offenbar durch den Begriff „Ergebnisse“ beeinflusst, der in der Formulierung von Niveau II enthalten ist und in der Erklärung wiederholt aufgegriffen wird.

5.2 Überarbeitung der Formulierungen

Aufgrund der im vorhergegangenen Abschnitt beschriebenen Verständnisschwierigkeiten wurden die Formulierungen selektiv überarbeitet. Krell und Krüger (2010) haben faktorenanalytisch drei Items für jedes Niveau der fünf Teilkompetenzen entwickelt. Diese Items dienten als Orientierung für die Optimierung der vorliegenden Formulierungen, insbesondere für die Wahl passender sprachlicher Alternativen. Vor allem in den Teilkompetenzen *Zweck*,

Testen und *Ändern* von Modellen stellten sich einzelne Begriffe als problematisch heraus (vgl. 5.1). Die Formulierungen wurden diesbezüglich überarbeitet und hinsichtlich ihrer Verständlichkeit erneut qualitativ untersucht ($N = 14$).

Tab. 2: Neue Formulierungen der Teilkompetenzen Alternative Modelle, Zweck, Testen und Ändern von Modellen. Die Platzhalter ([das Original] bzw. [das Modell]) werden in jeder Aufgabe durch das jeweilige Original oder Modell ersetzt.

<i>Ursprüngliche Formulierungen</i>		<u>Optimierte Formulierungen</u>
Alternative Modelle		
Es gibt verschiedene Modelle [des Originals], weil ...		
I	... die Modelle <i>im Erscheinungsbild unterschiedlich</i> sind.	... sich die Modelle <u>äußerlich unterscheiden</u> .
II	... bei <i>ihrer</i> Herstellung unterschiedliche Eigenschaften [des Originals] berücksichtigt wurden.	... bei <u>der Herstellung der Modelle</u> unterschiedliche Eigenschaften [des Originals] berücksichtigt wurden.
III	... mit <i>ihnen</i> unterschiedliche Vermutungen über [das Original] <i>erforscht</i> werden sollen.	... <u>mit den Modellen</u> unterschiedliche Vermutungen über [das Original] <u>gewonnen</u> werden sollen.
Zweck von Modellen		
[Das Modell] hat den Zweck, ...		
I	... [das Original] darzustellen.	... [das Original] <u>möglichst genau</u> darzustellen.
II	... <i>Verknüpfungen</i> zwischen Teilen [des Originals] zu <i>erläutern</i> <u>Zusammenhänge</u> zwischen Teilen [des Originals] zu <u>verdeutlichen</u> .
III	... <i>neue</i> Vermutungen über [das Original] zu <i>erforschen</i> <u>mit dem Modell</u> Vermutungen über [das Original] zu <u>gewinnen</u> .
Testen von Modellen		
[Das Modell] wird getestet, indem man ...		
I	... [das Modell] auf <i>Beschädigungen</i> kontrolliert.	... überprüft, ob [das Modell] <u>beschädigt</u> ist.
II	... nachprüft, ob <i>es</i> [dem Original] entspricht.	... überprüft, ob <u>das Modell</u> [dem Original] entspricht.
III	... <i>testet</i> , ob neue <i>Erwartungen</i> über [das Original] zutreffen.	... <u>mit dem Modell</u> überprüft, ob <u>Vermutungen</u> über [das Original] zutreffen.
Ändern von Modellen		
[Das Modell] wird verändert, um ...		
I	... <i>Beschädigungen</i> [am Modell] zu beheben.	... zu beheben, was [am Modell] <u>beschädigt</u> ist.
II	... neue <i>Ergebnisse</i> über [das Original] einzubeziehen.	... <u>in das Modell</u> neue <u>Erkenntnisse</u> über [das Original] einzubeziehen.
III	... <i>neue</i> Vermutungen über [das Original] <i>erforschen</i> zu können.	... <u>mit dem Modell</u> <u>weitere</u> Vermutungen über [das Original] <u>überprüfen</u> zu können.

Hierzu wurden den Probanden in einer Interviewsituation die bloßen Formulierungen (und keine FC-Aufgaben) vorgelegt, da das Erkenntnisinteresse nicht mehr auf den Denkprozessen während der Aufgabenbearbeitung lag. Die Probanden hatten die Aufgabe, die Formulierungen in ihren eigenen Worten zu erklären. Diese Erklärungen wurden, analog dem obigen Vorgehen, inhaltlich analysiert. Die inhaltlich validierten überarbeiteten Formulierungen sind in Tabelle 2 dargestellt.

5.3 Entscheidungsprozesse (Fragestellung 2)

In 21 von 43 Uminterpretationen (49 %) kommt es zu einer Umdeutung eines Niveaus, wenn nicht die eigene erste Präferenz zur Wahl steht. Es wird dabei in dem Paarvergleich, in dem die favorisierte Aussage nicht enthalten ist, ein Niveau ausgewählt und dies im Sinne der favorisierten Aussage umgedeutet. Tabelle 1 zeigt die Anzahl (A) der Umdeutungen (U), die auf das Fehlen der ersten Präferenz zurückzuführen sind. Besonders deutlich tritt dieses Phänomen bei der Teilkompetenz *Eigenschaften von Modellen* auf, hier sind sieben von acht Uminterpretationen durch dieses Phänomen erklärbar (Tab. 1).

5.4 Weitere Perspektiven (Fragestellung 3)

In der Teilkompetenz *Alternative Modelle* wurde eine Perspektive als Begründung für das präferierte Niveau (in zehn Fällen bei Niveau I, in drei Fällen bei Niveau II und in einem Fall bei Niveau III) eingenommen, die im Kompetenzmodell nicht beschrieben wird. Dabei wurde, trotz der Intention der Aufgabe, die alternativen Modelle nur für *ein* Original zu begründen, die Existenz alternativer Modelle auf unterschiedliche Originale bezogen. Exemplarisch hierzu Schüler 14G07: „Weil bei Leuten, die eine Herzkrankheit haben, ist dann das Herz kleiner oder unförmiger und deshalb gibt es unterschiedliche Formen [Modelle].“ Das Kompetenzmodell beschreibt mögliche Gründe für alternative Modelle eines Originals, womit diese Perspektive nicht durch eines der Niveaus abgebildet ist.

In der Teilkompetenz *Testen von Modellen* äußerten sich einige Schülerinnen und Schüler über „beschädigte Originale“ und machten damit keine Aussagen zum Testen von Modellen. In allen anderen Fällen ließen sich die Äußerungen der Schülerinnen und Schüler inhaltlich den Niveaus des Kompetenzmodells zuordnen.

6 Diskussion

Unter methodologischen Gesichtspunkten ist das beschriebene Vorgehen der inhaltlichen Validierung von FC-Aufgaben mit Hilfe des Lauten Denkens ein paralleles *mixed method*-Design (FOSCHT et al., 2007), da für die Beantwortung der gleichen Fragestellung sowohl quantitative als auch qualitative Daten generiert, ausgewertet und schließlich gegenübergestellt wurden. Die Gegenüberstellung beider Datentypen erfolgte hierbei aus der Perspektive einer gemeinsamen qualitativen Fragestellung, da inhaltliche Validität ein nur qualitativ prüfbares Konstrukt ist (KELLE & REITH, 2008).

Im Rahmen der Fragestellung 1 konnten einzelne Verständnis- und Interpretationsschwierigkeiten insbesondere beim Niveau III gefunden werden. In diesen Fällen wurde das Modell aus der Herstellungsperspektive betrachtet, das heißt, als Abbild der Realität. In Anbetracht des vermutlich dominierenden Einsatzes von Modellen in Schule als Medium (CRAWFORD & CULLIN, 2004; VAN DRIEL & VERLOOP, 2002), ist es nicht verwunderlich, dass Schülerinnen und Schüler vor allem diese Sicht auf Modelle vertreten und folglich Schwierigkeiten haben, Niveau III im Sinne der Aufgabe zu interpretieren. Das Problem ist dabei weniger ein mangelndes Verständnis einzelner Formulierungen. Denn es ist nicht überraschend, dass Lernende, deren Modellverständnis sich auf Niveau I oder II bewegt, die Formulierung von Niveau III nicht gänzlich verstehen. Problematisch ist aus der Sicht der Aufgabenkonstruktion und Diagnose, dass solche Probanden in einigen Fällen trotzdem Niveau III als erste Präferenz auswählen. Aus dieser Perspektive kann deshalb auch eine attraktivere Formulierung von Niveau I und/oder II zu einer besseren Passung zwischen tatsächlichem Modellverständnis und Ergebnis der Aufgabenbearbeitung beitragen. Insbesondere die Formulierungen der Niveaus II und III aller Teilkompetenzen (außer *Eigenschaften von Modellen*) wurden überarbeitet, um eine stärkere inhaltliche Abgrenzung zu erreichen (Tab. 2). Dabei ließen sich aus den Schwierigkeiten und Begründungen der Schülerinnen und Schüler bei einzelnen Begriffen Anhaltspunkt für eine selektive Optimierung der Formulierungen finden.

Mit Hilfe des qualitativen Zugangs konnte ein Einblick in das Bearbeitungsverhalten der FC-Aufgaben gewonnen werden (Fragestellung 2). Schülerinnen und Schüler identifizieren während der Aufgabenbearbeitung zwar eine ihrer Vorstellung entsprechende Formulierung und wählen diese zwei Mal aus, entscheiden sich dann aber beim Vergleich der beiden nicht präferierten Niveaus – der bei den FC-Aufgaben ebenfalls geleistet werden muss – gar nicht nachge-

ordnet für eine weitere Perspektive, sondern in 49 % der Fälle nur für das Niveau, das sie leichter im Sinne der eigenen Vorstellung uminterpretieren können. Die von COOMBS (1950) theoretisch beschriebenen Entscheidungsprozesse beim Bearbeiten von FC-Aufgaben konnten empirisch also nur teilweise bestätigt werden: Schülerinnen und Schüler vertreten zwar konsequent ihr individuelles Verständnis (ihre „individuell norm“), entscheiden allerdings zwischen den beiden übrigen Formulierungen oft nicht aufgrund der inhaltlichen Vorgaben, sondern der interpretierten Nähe zum eigenen Verständnis. Diese Entscheidungen können einzelne Begriffe hervorrufen, wenn deren Bedeutung nicht im Satzzusammenhang beachtet wird. Wegen der unsicheren Interpretation der Ergebnisse der FC-Aufgaben bezüglich der zweiten Präferenz erlauben sie somit in unserem Beispiel nur eingeschränkt, was von MCCLOY et al. (2005) als allgemeines Merkmal von FC-Aufgaben beschrieben wird: eine Positionierung der Vorstellung eines Lernenden auf dem zwischen den Niveaus I bis III aufgespannten Kontinuum (UPMEIER ZU BELZEN & KRÜGER, 2010). Die FC-Aufgaben ermöglichen aber, das Niveau zu identifizieren, das den individuellen Vorstellungen am stärksten entspricht.

In der Teilkompetenz *Alternative Modelle* konnte eine Perspektive identifiziert werden, die nicht durch die Niveaus des Kompetenzmodells abgedeckt ist (Fragestellung 3). Diese wurde bereits von JUSTI und GILBERT (2003) sowie GRÜNKORN et al. (2011) beschrieben: Für viele Schülerinnen und Schüler ist die wichtigste Begründung für alternative Modelle ein Unterschied zwischen Originalen. Dies widerspricht der Aufgabenintention, Gründe für verschiedene Modelle *eines* Originals zu beschreiben. Diese Vorstellung ergibt sich aus der Überzeugung, dass Modelle Kopien der Realität sind. Folglich werden von vielen Schülerinnen und Schülern die angebotenen Formulierungen in der Teilkompetenz *Alternative Modelle* auch aus dieser Perspektive betrachtet und (um-)gedeutet. Eine pauschale Einordnung dieser Vorstellung im Sinne eines basalen „Niveau 0“ im Kompetenzmodell würde eventuell zu kurz greifen, da auch mit Bezug auf mehrere Originale eine Herstellungs- oder Anwendungsperspektive ausgedrückt werden kann. Alternative Modelle können der Darstellung oder der Erforschung unterschiedlicher Originale dienen. Für die Diagnose des Modellverständnisses von Lernenden kann in die Aufgaben eine Vorauswahl im Sinne eines adaptiven Testens implementiert werden. Dann können nur diejenigen Probanden nach unterschiedlichen Gründen für alternative Modelle gefragt werden, die der Vorstellung, dass es nur ein Modell zu einem Original gibt, nicht zustimmen. Andernfalls würden die Aufgaben nur eine mangelhafte Varianzaufklärung erlauben, da eine „relevante Kausalbedingung des untersuchten Handelns nicht empirisch erfasst“ wird (KELLE, 2008, S. 230).

7 Fazit und Ausblick

Durch den qualitativen Zugang konnten wichtige Hinweise dafür gewonnen werden, dass – und an welchen Stellen – die auf der Basis des Kompetenzmodells der Modellkompetenz konstruierten Formulierungen (deduktive bzw. rationale Testkonstruktion; HARTIG & JUDE, 2007) Verständnisschwierigkeiten bei Schülerinnen und Schülern hervorrufen. Die Formulierungen wurden auf dieser Grundlage selektiv optimiert und erneut evaluiert.

Es konnte darüber hinaus eine inhaltliche Perspektive, die bereits von JUSTI und GILBERT (2003) und GRÜNKORN et al. (2011) beschrieben wurde, auch im Rahmen der vorliegenden Untersuchung nachgewiesen werden – die *uniqueness* von Modellen. Schließlich konnte gezeigt werden, dass die Ergebnisse der FC-Aufgaben weniger als Positionierung auf einem Kontinuum (MCCLOY et al., 2005), denn als Präferenz für eines der drei Niveaus interpretiert werden sollten. Diese Erkenntnisse wären im Rahmen einer rein quantitativen Untersuchung verschlossen geblieben, wodurch die Sinnhaftigkeit einer qualitativen Ergänzung auch (bzw. insbesondere) bei deduktiv angelegten Forschungsprojekten untermauert wird. Die validierten FC-Aufgaben werden im Rahmen einer quantitativen Untersuchung eingesetzt, um mit Hilfe statistischer Analysen Rückschlüsse auf die im Kompetenzmodell postulierte Struktur von Modellkompetenz und auf den Einfluss des konkreten Aufgabenstamms auf das Ergebnis der Aufgabenbearbeitung zu gewinnen (KRELL & KRÜGER, 2011).

Zitierte Literatur

- ARN, W. (1990). *Phänomene zwischen Natur und Technik*. Zürich & Wiesbaden: Orell Füssli.
- AUFSCHNAITER, C. von & Rogge, C. (2010). Wie lassen sich Verläufe der Entwicklung von Kompetenz modellieren? *ZfDN*, 16, 95-114.
- BÖCKENHOLT, U. (2004). Comparative judgment as an alternative to ratings: Identifying the scale origin. *Psychological Methods*, 9, 453-465.
- BORTZ, J., LIENERT, G. & BOEHNKE, K. (2008). *Verteilungsfreie Methoden in der Biostatistik*. Heidelberg: Springer.
- BROWN, A. & BARTRAM, D. (2009, April). Doing less but getting more: Improving forced-choice measures with IRT. *Paper presented at the 24th Annual conference of the Society for Industrial and Organizational Psychology*, New Orleans, USA, 28 Seiten.
- BROWN, A. & MAYDEU-OLIVARES, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 3, 460-502.
- CAMPBELL, N. & REECE, J. (2011). *Campbell Biologie. Gymnasiale Oberstufe*. München: Pearson.
- CHAN, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, 30, 99-121.
- CHITTLEBOROUGH, G., TREAGUST, D., MAMIALA, T. & MOCERINO, M. (2005). Students' perceptions of the role of models in the process of science and in the process of learning. *Research in Science and Technological Education*, 2, 195-212.

- CLANCY, K. & GARSEN, R. (1970). Why some scales predict better. *Journal of Advertising Research*, 10, 33-38.
- COOMBS, C. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57, 145-158.
- CRAWFORD, B. & CULLIN, M. (2004). Supporting prospective teachers' conceptions of modeling in science. *IJSE*, 26, 1379-1401.
- EISMA, J. (2012). *Flow visualization and force measurements on a flapping-wing MAV DelFly II in forward flight configuration*. Master of Science Thesis, Delft University of Technology.
- ERICSSON, K. & SIMON, H. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.
- FOSCHT, T., ANGERER, T. & SWOBODA, B. (2007). Mixed Methods. Systematisierung von Untersuchungsdesigns. In R. Buber & H. H. Holzmüller (Hrsg.), *Qualitative Marktforschung* (S. 247-259). Wiesbaden: Gabler.
- FRANKE, G. (2005). *Facetten der Kompetenzentwicklung*. Bielefeld: Bertelsmann.
- GILBERT, J., BOULTER, C. & EIMER, R. (2000). Positioning models in science education and in design and technology education. In J. K. Gilbert & C. J. Boulter (Hrsg.), *Developing models in science education* (S. 3-17). Dordrecht [u.a.]: Kluwer Academic.
- GROPENGIEßER, H. (2005). Qualitative Inhaltsanalyse in der Lehr-Lernforschung. In P. Mayring (Hrsg.), *Die Praxis der Qualitativen Inhaltsanalyse* (S. 172-189). Weinheim & Basel: Beltz.
- GROSSLIGHT, L., UNGER, J., JAY, E. & SMITH, C. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *JRST*, 28, 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. Carvalho (Hrsg.), *Authenticity in biology education: Benefits and challenges. A selection of papers presented at the 8th Conference of European Researchers in Didactics of Biology (ERIDOB) Braga, Portugal*, 53-65.
- HARRISON, A. & TREAGUST, D. (2000). A typology of school science models. *IJSE*, 22, 1011-1026.
- HARTIG, J. & JUDE, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik: Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung* (S. 17-36). Bonn & Berlin: BMBF.
- HARTIG, J., KLIEME, E. & LEUTNER, D. (Hrsg.) (2008). *Assessment of competencies in educational contexts*. Göttingen: Hogrefe.
- HICKS, L. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.
- JUSTI, R. & GILBERT, J. (2003). Teacher's views on the nature of models. *IJSE*, 25, 1369-1386.
- KATTMANN, U. (2006). Modelle. In H. Gropengießer & U. Kattmann (Hrsg.), *Fachdidaktik Biologie* (S. 330-339). Köln: Aulis.
- KELLE, U. (2008). *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung*. Wiesbaden: VS.
- KELLE, U. & REITH, F. (2008): Empirische Forschungsmethoden. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion* (S. 39-75). Wiesbaden: VS.
- KELLE, U. & KLUGE, S. (2010). *Vom Einzelfall zum Typus*. Wiesbaden: VS.
- KLEICKMANN, T., HARDY, I., MÖLLER, K., POLLMEIER, J., TRÖBST, S. & BEINBRECH, C. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion. *ZfDN*, 16, 265-283.
- KRELL, M. & KRÜGER, D. (2010). Diagnose von Modellkompetenz. Deduktive Konstruktion und Selektion von geschlossenen Items. *Erkenntnisweg Biologiedidaktik*, 9, 23-37.
- KRELL, M. & KRÜGER, D. (2011): Forced Choice-Aufgaben zur Evaluation von Modellkompetenz im Biologieunterricht – Empirische Überprüfung konstrukt- und merkmalsbezogener Teilkompetenzen. *Erkenntnisweg Biologiedidaktik*, 10, 53-67.
- KUNTER, M. & KLUSMANN, U. (2010). Kompetenzmessung bei Lehrkräften – Methodische Herausforderungen. *Unterrichtswissenschaft*, 38, 68-86.
- MAHR, B. (2008). Ein Modell des Modellseins. In U. Dirks & E. Knobloch (Hrsg.), *Modelle* (S. 187-218). Frankfurt am Main: Peter Lang.
- MAYRING, P. (2000). Qualitative Inhaltsanalyse. *Forum Qualitative Sozialforschung*, 1, 1-10.

- MCCLOY, R., HEGGESTAD, E. & REEVE, C. (2005). A silk purse from the sow's ear: Retrieving normative information from multidimensional forced-choice items. *Organizational Research Methods*, 8, 222-248.
- PEIRCE, C. (1967). *Schriften zum Pragmatismus und Pragmatizismus*. Herausgegeben von Karl-Otto Apel, Frankfurt am Main: Suhrkamp.
- RUPP, A., LEUCHT, M. & HARTUNG, R. (2006). ‚Die Kompetenzbrille aufsetzen‘. *Unterrichtswissenschaft*, 34, 195-219.
- SCHECKER, H. & PARCHMANN, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *ZfDN*, 12, 45-66.
- SCHOLL, A. (2009). *Die Befragung*. Konstanz: UVK.
- TASHAKKORI, A. & TEDDLIE, C. (2003). *Handbook of mixed methods in social and behavioral research*. Los Angeles: Sage.
- TERZER, E. & UPMEIER ZU BELZEN, A. (2010). Modellkompetenz im Biologieunterricht – Empirische Überprüfung eines Kompetenzmodells mit Multiple Choice-Items. In M. Gräsel-Zikuda, C. Rohlf, & T. Seidel (Hrsg.), *74. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung* (S. 234). Jena.
- TRIER, U. & UPMEIER ZU BELZEN, A. (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. & KRÜGER, D. (2010). Modellkompetenz im Biologieunterricht. *ZfDN*, 16, 41-57.
- VAN DRIEL, J. & VERLOOP, N. (2002). Experienced teacher's knowledge of teaching and learning of models and modeling in science education. *IJSE*, 24, 1255-1277.

