

Vermittlung von Modellkompetenz in den Unterrichtsfächern Biologie und Chemie

Modellierung, Validierung und Messung
Professioneller Unterrichtswahrnehmung zukünftiger Lehrkräfte mithilfe
eines Vignettentests

Von der Pädagogischen Hochschule Heidelberg zur Erlangung des Grades eines
Doktors der Philosophie (Dr. phil.) genehmigte Dissertation von

Benjamin Johannes Tempel

aus

Bietigheim-Bissingen

(Geburtsort)

2017

(Jahr der Annahme der Dissertation)

Erstgutachter: Prof. Dr. Markus Wilhelm

Zweitgutachter: Prof. Dr. Markus Rehm

Fach: Biologie- und Chemiedidaktik

Tag der mündlichen Prüfung: 23.03.2017

Danksagung

Diese Dissertation ist das Produkt eines umfangreichen wissenschaftlichen Vorhabens. Zu den Menschen, bei denen ich mich besonders für die Unterstützung bedanken möchte, zählen die Lehrbeauftragten der Staatlichen Seminare für Didaktik und Lehrerbildung (Realschulen) in Karlsruhe und Ludwigsburg. Durch ihre Rückmeldung haben sie zum einen erheblich zu einer höheren Passung des Tests zur Unterrichtsrealität beigetragen, zum anderen hilfreiche Hinweise zur Itemgestaltung und -formulierung eingebracht. Danken möchte ich auch allen Personen aus Deutschland, Österreich und der Schweiz, die aufgrund ihrer Lehrexpertise an der ersten und zweiten quantitativen Expertinnen- und Expertenbefragung teilnahmen und so die Bildung des Referenzwertes ermöglichten. Ebenfalls ganz herzlich danken möchte ich den Hilfswissenschaftlerinnen und Hilfswissenschaftlern, die mir bei der Beschaffung der ca. 600 Artikel im Rahmen des systematischen Literaturreviews halfen.

Namentlich will ich Prof. Dr. Christoph Randler für seine pragmatische Sicht und Beratung sowie Dipl.-Päd. Stella Ekler für die gute Zusammenarbeit danken. Herrn Dr. Ernst Lexen danke ich für seine inspirierende Freude am Programmieren mit R und seine fundierte mathematische Argumentation, durch die ich meine Ergebnisse präzisieren konnte. Meinem Erstbetreuer Herrn Prof. Dr. Markus Wilhelm danke ich herzlich für seine ansteckende Freude an der Naturwissenschaft und seine fruchtbare, den Blick weitende Sicht, die beispielsweise zur Berücksichtigung des Reasoning-Konstrukts führte. Herzlich will ich auch meinem Zweitbetreuer Prof. Dr. Markus Rehm für seine seit meiner Masterarbeit 2012 stets offene Tür danken und für die Gespräche, die mich Schritt für Schritt zu einer fundierten Theoriekenntnis führten.

Meinen Eltern und Schwestern danke ich, weil sie zu meinem wissenschaftlichen Interesse beitrugen. Zuletzt geht mein Dank an meine liebe Frau Sarah, die immer wieder meine Gedanken zurechtrückte und Freiräume schuf, sowie an meine Söhne Johannes und Juri, durch die ich mich immer wieder auf das Wesentliche besinnen konnte. Ohne ihre Unterstützung wäre diese Promotion nicht möglich gewesen.

Inhaltsverzeichnis

Tabellenverzeichnis.....	7
Abbildungsverzeichnis	8
Teil I: Theorie.....	13
1 Einleitung.....	13
2 Theoretische Ansätze und Methoden zur Erhebung von Lehrkompetenz	16
2.1 Lehrkompetenzen	18
2.1.1 Kategorien des Professionswissens nach Shulman	19
2.1.2 Kompetenzmodell von COACTIV.....	21
2.2 Quantitative Studien zur Lehrkompetenzstruktur	22
2.2.1 COACTIV	22
2.2.2 MT21.....	24
2.3 Vignettentests: ökologisch generalisierbare Kompetenzmessung	25
2.3.1 Der Vignettentest als Forschungsinstrument.....	25
2.3.2 Vignettentests im Vergleich	27
3 Professionelle Unterrichtswahrnehmung des Modelleinsatzes im Unterricht	31
3.1 Modelle und Modellkompetenz im Chemie- und Biologieunterricht	31
3.2 Professionelle Unterrichtswahrnehmung	38
3.2.1 Professionelle Unterrichtswahrnehmung als spezielle Kompetenz	38
3.2.2 Reasoning	40
3.2.3 Noticing.....	42
A) Modelle adäquat einsetzen können	48
B) Modelle (weiter)entwickeln können.....	49
C) Präkonzepte erkennen können.....	50
D) Conceptual Change	51
E) Modellmethode.....	51
F) Wissen um den Unterschied verschiedener Modelltypen.....	52

G) Mit diskontinuierlichen Modellen und Realmodellen umgehen können	52
H) Mit Teilchen und Atomen umgehen können.....	53
I) Modellieren vermitteln können	54
3.3 Zusammenfassung und Schlussfolgerungen aus dem Theorieteil.....	56
3.4 Forschungsfragen und Hypothesen	57
Teil II: Methoden, empirisch-statistisches Vorgehen und Ergebnisse.....	59
4 Methoden	59
4.1 Vignettenentwicklung	59
4.2 Kovariaten	62
4.3 Validierungsschritte	65
4.4 Bepunktung	68
5 Statistisches Vorgehen	71
5.1 Stichprobenbeschreibung	71
5.2 Statistische Analyseverfahren und Software.....	71
5.3 Umgang mit fehlenden Werten	72
6 Ergebnisse	78
6.1 Testbeschreibung und Reliabilität.....	78
6.2 Validitätshinweise	79
6.2.1 Kovariaten als Hinweise auf Inhalts- und Kriteriumsvalidität.....	80
6.2.2 Soziodemografische Hinweise auf Inhalts- und Kriteriumsvalidität	84
6.2.3 Test- und Itemschwierigkeiten	89
6.3 Empirische Überprüfung der Struktur des Noticings und des Reasonings zur Vermittlung von Modellkompetenz.....	96
6.3.1 Empirische Überprüfung der Struktur des Reasonings	96
6.3.2 Empirische Überprüfung der Struktur des Noticings.....	98
6.4 Exploratorische Faktorenanalysen	102
7 Diskussion, Limitation, Ausblick und Zusammenfassung	106
7.1 Diskussion	106

7.2	Limitation	115
7.3	Ausblick	116
7.4	Zusammenfassung	119
8	Literaturverzeichnis	126
9	Anhang	150
9.1	Eidesstattliche Versicherung	150
9.2	Tabellen zum systematischen Literaturreview	151
9.2.1	Überblick über Art des Artikels, Modelldefinition, Thematisierung von Modellen und Herkunftsland	151
9.2.2	Tabelle mit zentralen Forschungsergebnissen, Forschungsfragen, Auswertungsmethoden, Stichprobengröße und Population	171
9.2.3	Tabelle mit Zentralbegriffen und den zugeordneten Unterrichtsfächern	225
9.2.4	Tabelle zu den quantitativen Studien und deren statistischen Schwächen, falls keine Vergleichsgruppe, keine Stichprobe und keine Randomisierungs- oder Auswertungverfahren genannt wurden	229

Tabellenverzeichnis

Tabelle 1: Überblick über Modelltypen im Biologieunterricht.....	34
Tabelle 2: Ergänzender Überblick über Modelltypen im Chemieunterricht.....	36
Tabelle 3: Gegenüberstellung bisheriger Reasoning-Facetten.....	40
Tabelle 4: Zentrale soziodemografische Werte des Chemie- und Biologietests.....	71
Tabelle 5: Perzentilauswertung des Chemiedatensatzes	74
Tabelle 6: Perzentilauswertung des Biologiedatensatzes	75
Tabelle 7: Mplus-basierte Berechnung mit Paarvergleichen (vgl. 4.4)	75
Tabelle 8: Mplus-basierte Berechnung mit Paarvergleichen (vgl. 4.4)	75
Tabelle 9: Übersicht über die durchschnittlichen Testergebnisse.....	78
Tabelle 10: Reliabilität der entwickelten Tests	78
Tabelle 11: Reliabilitätswerte der Kovariaten	80
Tabelle 12: Manifeste Korrelation zwischen Chemietestscore und Kovariaten bzw. deren Subskalen	81
Tabelle 13: Manifeste Korrelation zwischen den Biologiescores und Kovariaten bzw. deren Subskalen	82
Tabelle 14: Manifeste Korrelationen zwischen Alter und Abiturnote und den Chemie- und Biologiescores	85
Tabelle 15: Mplus-basierte, geominrotierte Faktorladungen einer exploratorischen Faktorenanalyse mit dem Chemiedatensatz $NA < 50 \% = 0$	103
Tabelle 16: Mplus-basierte, geominrotierte Faktorladungen einer exploratorischen Faktorenanalyse mit dem Biologiedatensatz $NA < 50 \% = 0$	104
Tabelle 17: Überblick über die Passung verschiedener konfirmatorischer Faktorenanalysen und Validitätshinweise in Abhängigkeit von den Datensätzen.	122

Abbildungsverzeichnis

Abbildung 1: PID-Modell nach Blömeke et al. (2015, S. 7).....	17
Abbildung 2: Interconnected Model of Teacher Professional Growth – IMTPG (Clarke & Hollingsworth, 2002, S. 948).....	19
Abbildung 3: COACTIV-Kompetenzmodell nach Baumert und Kunter (2011, S. 32), eigene Darstellung.....	22
Abbildung 4: Strukturdiagramm zur Operationalisierung der <i>Elemente zur Vermittlung von Modellkompetenz</i> (Noticing) im Biologieunterricht. In Klammern findet sich die Anzahl der Artikel, die den jeweiligen Aspekt benennen. Hellgrau sind all jene Facetten, die bei der Entwicklung des Vignettentests unberücksichtigt bleiben mussten, eigene Darstellung.....	45
Abbildung 5: Strukturdiagramm zur Operationalisierung der <i>Elemente zur Vermittlung von Modellkompetenz</i> (Noticing) im Chemieunterricht. In Klammern findet sich die Anzahl der Artikel, die den jeweiligen Aspekt benennen. Hellgrau sind all jene Facetten, die bei der Entwicklung des Vignettentests nicht berücksichtigt werden konnten, eigene Darstellung.....	47
Abbildung 6: Modellierungskreislauf nach Blum (2006, S. 9).....	54
Abbildung 7: Modellierungskreislauf in der Chemiedidaktik nach Johnstone (1991, S. 81); adaptierte Übersetzung: B. T, eigene Darstellung.....	55
Abbildung 8: Biologie-Textvignette zur Noticing-Facette <i>Präkonzepte</i> mit typischem Aufbau und Itemauswahl, eigene Darstellung.....	61
Abbildung 9: Chemie-Textvignette zur Noticing-Facette <i>Präkonzepte</i> mit typischem Aufbau und Itemauswahl, eigene Darstellung.....	62
Abbildung 10: Beispiel eines Teils einer Aufgabe aus dem Wissenstests über Modellunterscheidung im Kontext des Wissens über die Natur der Naturwissenschaften (Nature of Science, NoS) im Biologieunterricht, umgesetzt in Sosci-Survey (Leiner & Leiner, 2015).....	63
Abbildung 11: Antwortverteilung der Expertinnen und Experten: Aufgrund eindeutiger Antworten konnte das Item E9V23I05 bei der weiteren Auswertung berücksichtigt werden, Item E9V24I03 hingegen nicht, eigene Darstellung.....	68
Abbildung 12: Vergleich zweier Items hinsichtlich der Verteilung der Antworten der Probandinnen und Probanden, eigene Darstellung.....	69

- Abbildung 13: Hypothetisches Beispiel für die Bepunktung des Paarvergleichs (PV): Die Modalwerte des Expertinnen- Expertenratings (X_{Ex}) werden je Itempaar miteinander verglichen. Proband 1 (P1) hat im Gegensatz zum aggregierten Expertinnen- und Expertenurteil nicht erkannt, dass die in Item 2 getroffene Aussage besser ist als die in Item 1 getroffene Aussage. Da sich die Modalwerte des Expertinnen- und Expertenratings jedoch nur im Betrag 1 voneinander unterscheiden, wird diese Relation nicht berücksichtigt. Entscheidend sind nur der PV zwischen Item 1 und Item 3 sowie Item 2 und Item 3. Proband 1 hat erkannt, dass Item 1 und Item 2 besser als Item 3 sind und bekommt hierfür jeweils einen Punkt. Im Gegensatz dazu erhält Proband 2 (P2) für beide Relationen je 0 Punkte, da er die in Item 3 getroffene Aussage für die beste Lösung hält. 69
- Abbildung 14: Querschnittliche Entwicklung des Chemietestscores über die gruppierten Unterrichtspraktikumsstunden. Es zeigen sich signifikante ($p = .020$, $F(2, 246) = 3.94$) Unterschiede zwischen bis zu acht Stunden und mehr als 50 Stunden Unterrichtserfahrung, eigene Darstellung. 86
- Abbildung 15: Querschnittliche Entwicklung des Chemietestscores über gruppierte Fachsemester. Es zeigen sich marginal signifikante (\dagger : $p = .095$, $F(3, 268) = 3.46$) Unterschiede zwischen 1.-3. Fachsemester und 4.-6. Fachsemester sowie signifikante ($p = .013$, $F(3, 268) = 3.46$) Unterschiede zwischen 1.-3. Fachsemester und 7.-8. Fachsemester, eigene Darstellung. 87
- Abbildung 16: Querschnittliche Entwicklung des Biologiegesamttestscores über die Fachsemester. Es zeigen sich signifikante ($p = .034$, $F(8, 205) = 1.94$) Unterschiede zwischen erstem Fachsemester und Referendariat, eigene Darstellung. 88
- Abbildung 17: Querschnittliche Entwicklung des Biologiegesamttestscores über die gruppierten Fachsemester. Es zeigen sich signifikante ($p = .048$, $F(3, 210) = 5.05$) Unterschiede zwischen dem ersten bis dritten Fachsemester und dem siebten bis achten Semester und hochsignifikante ($p = .006$, $F(3, 210) = 5.05$) Unterschiede zwischen dem ersten bis dritten Fachsemester und dem Referendariat, eigene Darstellung. 88
- Abbildung 18: Biologie-Vignettentest und -NoS-Test (Datensatz $NA = 0$): Anzahl der richtig gelösten (Pseudo-) Items bzw. Paarvergleiche. Die gestrichelte Linie entspricht 50 % der Probandinnen und Probanden ($N = 188$), eigene Darstellung. 89
- Abbildung 19: Testschwierigkeit des Biologievignettentests in Abhängigkeit von der Itemschwierigkeit (Datensatz: $NA = 0$, $N = 376$, ohne Modellvergleichs-/NoS-Items),

- berechnet mit dem R-Package mirt (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. P steht für den Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung. 90
- Abbildung 20: Testschwierigkeit des Biologievignettentests in Abhängigkeit von der Itemschwierigkeit (Datensatz: $NA < 50 \% = 0$, $N = 348$), berechnet mit dem R-Package mirt (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. MN steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung 91
- Abbildung 21: Testschwierigkeit des Biologievignettentests in Abhängigkeit von der Itemschwierigkeit (Datensatz: $NA < 50 \% = imp.$ $N = 348$), berechnet mit dem R-Package mirt (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. MN steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung..... 92
- Abbildung 22: Testschwierigkeit des Chemie vignettentests (Datensatz $NA = 0$) in Abhängigkeit von der Itemschwierigkeit, berechnet mit dem R-Package mirt (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. P steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung..... 93
- Abbildung 23: Testschwierigkeit des Chemie vignettentests (Datensatz: $NA < 50 = 0$) in Abhängigkeit von der Itemschwierigkeit, berechnet mit dem R-Package mirt (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. M steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung..... 94

- Abbildung 24: Testschwierigkeit des Chemie vignettentests (Datensatz: $NA < 50 = imp.$) in Abhängigkeit von der Itemschwierigkeit, berechnet mit dem R-Package mirt (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. M steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung..... 95
- Abbildung 25: Reasoning CFA des Chemietests. Alle Itemscores fielen zu einem Generalfaktor zusammen. Ein fittendes Modell hinsichtlich RMSEA und CFI konnte über Fitindices nicht bestätigt werden. *Abkürzungen:* PVS: Paarvergleichssumme; B: Bewertungsmöglichkeiten auswählen können; D: Diagnosemöglichkeiten auswählen können; H: Handlungsalternativen auswählen können; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index, eigene Darstellung..... 97
- Abbildung 26: Reasoning CFA des Biologietests. Ein fittendes Modell hinsichtlich RMSEA und CFI wurde über Fitindices bestätigt. Die starke Korrelation zwischen Diagnosemöglichkeiten/Handlungsalternativen und Bewertungsmöglichkeiten ist signifikant. *Abkürzungen:* PVS: Paarvergleichssumme; B: Bewertungsmöglichkeiten auswählen können; D: Diagnosemöglichkeiten auswählen können; H: Handlungsalternativen auswählen können; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index, eigene Darstellung 98
- Abbildung 27: Modellkompeten-Noticing CFA Chemie. Ein fittendes Modell hinsichtlich RMSEA und CFI wurde gefunden, der TLI ist knapp nicht akzeptabel. *Abkürzungen:* PVS: Paarvergleichssumme; B: Bewertungsmöglichkeiten auswählen können; D: Diagnosemöglichkeiten auswählen können; H: Handlungsalternativen auswählen können; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index. Bis auf die Kovarianz zwischen *Modelle adäquat einsetzen können* und *Unterschiede beim Einsatz von Teilchen- und Atomvorstellungen erkennen können* ($p = 0.537$) sind die latenten Korrelationen signifikant, eigene Darstellung. 99
- Abbildung 28: Noticing-CFA Biologie. Ein akzeptables bis gutes Modell hinsichtlich RMSEA sowie CFI und TLI wurde gefunden. Alle latenten Korrelationen werden signifikant. *Abkürzungen:* PVS: Paarvergleichssumme; B: Bewertungsmöglichkeiten

auswählen können; D: Diagnosemöglichkeiten auswählen können; H: Handlungsalternativen auswählen können; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index, eigene Darstellung	101
Abbildung 29: Studienergebnisse mit besonderer Relevanz für die Unterrichtspraxis, eigene Darstellung	112

Teil I: Theorie

1 Einleitung

All models are wrong, but some are useful (vgl. Box, 1976).

Naturwissenschaftliche Modelle sind zentrale Arbeits- und Denkweisen in der Biologie und Chemie (vgl. Upmeier zu Belzen & Krüger, 2010, S. 42). Fähigkeiten im Vermitteln¹ von Modellkompetenz sind somit auch ein wichtiger Teil der Lehrkompetenz in den Schulfächern Biologie und Chemie. Bisherige Studien zu Modellen im Unterricht untersuchen zumeist das Wissen über Modelle (Justi & van Driel, 2005a, 2005b) von Lehrkräften und Schülerinnen und Schülern. Es bestehen jedoch Zweifel in der fachdidaktischen Forschung, ob das Wissen über Modelle zentral ist für die Vermittlung von naturwissenschaftlicher Modellkompetenz (vgl. Upmeier zu Belzen & Krüger, 2010). Studien zu Modellwissen und zur quantitativ angelegten fachdidaktischen Forschung, die sich an deklarativem² Wissen orientieren, sind in die Kritik geraten. Es wird bezweifelt, ob solche Tests dem Anspruch auf Abbildung von Lehrkompetenz gerecht werden (Kromrey & Renfrow, 1991; McClelland, 1973).

Um diesem Anspruch zu genügen, wäre eine direkte Erfassung des Unterrichts durch teilnehmende Unterrichtsbeobachtung oder eine Videostudie naheliegend. Damit würde dem Bedürfnis nach Erfassung von lernwirksamen Tiefenstrukturen von Unterricht (Kunter & Trautwein, 2013, S. 76) entsprochen. Doch zum einen ist Unterrichtsbeobachtung, die über Sichtstrukturen z. B. der Modellvermittlungsmethode hinausgeht, empirisch nur schwer zu erfassen (vgl. Seidel et al., 2006, 815 f.), zum anderen besteht ein Anspruch auf verallgemeinerbare Aussagen. Hierzu sind jedoch nur Forschungsinstrumente denkbar, die reliabel und in großen Stichproben einsetzbar sind. Die direkte Erfassung von Lehrkräftenhandeln im Biologie- und Chemieunterricht kann dieser Ambition aus forschungsökonomischen Gründen nicht gerecht werden.

Deshalb wird stellvertretend nach handlungsrelevanten und dennoch in großen Stichproben einsetzbaren validen Forschungsinstrumenten gesucht. Dieser Ambition wird hier indirekt

¹ Unter *Vermittlung* wird im Folgenden ein Lernangebot verstanden, dass mit vergleichbar hoher Wahrscheinlichkeit von Schülerinnen und Schülern genutzt wird (vgl. das Angebots-Nutzungs-Modell nach Fend (2001)).

² Nach Gruber (2008) wird unter deklarativem Wissen leicht überprüfbares Faktenwissen verstanden.

über das Konstrukt *Professionelle Unterrichtswahrnehmung* (PU, vgl. Goodwin, 1994) begegnet. Dadurch sollen auf dem Umweg über die Interpretation der Wahrnehmung von Unterricht Rückschlüsse auf die Fähigkeit, Modellkompetenz zu vermitteln, gezogen werden. Die Dissertation fokussiert daher auf die Entwicklung eines Vignettentests (vgl. Stecher et al., 2006), d. h. auf die Entwicklung eines quantitativen Testverfahrens, das Unterrichtssituationen in Form von Videos und Texten präsentiert (vgl. 2.3.1, S. 25). Hierbei sollen künftige Lehrkräfte auf einer sechsstufigen Skala Items beantworten. Studien mit dem Testformat Vignettentest weisen teilweise positiv-signifikante Korrelationen zwischen Testscore und Unterrichtsqualität bzw. Schülerinnen- und Schülerleistung auf (Kersting, Givvin, Thompson, Santagata & Stigler, 2012).

Ziel der Dissertation ist wie bei allen anderen EKoL³-Teilprojekten die Entwicklung, Modellierung, Validierung und Auswertung je eines Vignettentests. Die in dieser Studie entstandenen zwei Vignettentests für den Biologie- und Chemieunterricht sollen Rückschlüsse auf die Professionelle Unterrichtswahrnehmung (als Überblick: Meschede, 2014) zur Vermittlung von Modellkompetenz in den jeweiligen Domänen ermöglichen. Unter Professioneller Unterrichtswahrnehmung wird sowohl *Noticing* (selektive Aufmerksamkeit, vgl. Stürmer & Seidel, 2015) als auch *Reasoning* (wissensbasierte Schlussfolgerungen, vgl. Stürmer & Seidel, 2015) verstanden. Beide Aspekte wurden bei der Entwicklung der Tests berücksichtigt und bei der Durchführung an künftigen Lehrkräften empirisch überprüft. Das hier vorgestellte Forschungsinstrument wurde mehrfach validiert. Es werden Hinweise auf Inhalts- und Kriteriumsvalidität präsentiert.

Der Aufbau der Arbeit folgt zunächst dem Ziel, (Lehr-)Kompetenzen in der Chemie- und Biologiedidaktik sowie relevante Studien zur Erhebung von Lehrkompetenz vorzustellen, um anschließend im Speziellen auf Modelle und Modellkompetenzen einzugehen. Es folgen definitorische Bestimmungen von Professioneller Unterrichtswahrnehmung, *Noticing* und *Reasoning*. Die Forschungsfragen und Hypothesen greifen das theoretisch Erarbeitete auf und fokussieren auf das Ziel dieser Arbeit: die Entwicklung, Modellierung und Validierung je eines Vignettentests zur Vermittlung von Modellkompetenz in den Unterrichtsfächern Chemie und Biologie. Dabei wird geprüft, ob sich die aus einem systematischen Literaturreview gewonnenen Erkenntnisse über die selektive Wahrnehmung bzw. das *Noticing* der *Elemente der Vermittlung von Modellkompetenz* in den empirisch erhobenen Daten wiederfinden. Zu-

³ Diese Dissertation entstand im Rahmen des strukturierten Forschungs- und Nachwuchskollegs EKoL (Effektive Kompetenzdiagnose in der Lehrerbildung).

dem wird anhand der empirischen Daten geprüft, ob sich in den Fragen zu den Itembatterien, die die Probandinnen und Probanden der hier entwickelten Vignettentests gestellt bekommen, die enthaltene Reasoning-Struktur wiederfinden lässt. Anschließend werden die Entwicklung und die Validierung des Instruments beschrieben. Darauf folgt die Beschreibung der Stichprobe, des Analyseverfahrens und des Umgangs mit fehlenden Werten. Der Ergebnisteil der Arbeit gibt Hinweise auf die Validität der Vignettentests. Hierzu werden zunächst die Reliabilitätswerte dargestellt. Anschließend folgen Hinweise zur Validität, die über Kovariaten, Soziodemografie und Test- bzw. Itemschwierigkeiten gewonnen wurden. Hierauf folgt die Beantwortung der aus dem Ziel der Testentwicklung resultierenden Frage nach der Passung der empirischen Daten zum theoriebasierten Noticing- und Reasoning-Konstrukt. Dies erfolgt jeweils gesondert für den Chemie- und Biologietest. Für den Chemietest wird sodann eine alternative, explorativ entwickelte Struktur vorgeschlagen. Die Studie endet mit der Diskussion und der Zusammenfassung der Ergebnisse. Hierbei werden auch für die Unterrichtspraxis und die Aus- und Weiterbildung von Lehrkräften relevante Implikationen dargestellt. Außerdem soll das Schlusskapitel einen Ausblick für zukünftige Forschung im Bereich der Vermittlung von Modellkompetenz bieten.

2 Theoretische Ansätze und Methoden zur Erhebung von Lehrkompetenz

Empirisch-quantitative Tests können das Ziel verfolgen, Leistung zu messen. Welcher Art diese Leistung ist, unterscheidet sich. Denkbar sind Tests z. B. zu Wissen und Kompetenz. Deklarative Wissenstests (siehe Fußnote 2, S. 13) sind weit verbreitet, stehen jedoch seit Langem in der Kritik, beispielsweise wenn sie zu Assessments von (künftigen) Lehrkräften eingesetzt werden (vgl. Kromrey & Renfrow, 1991; McClelland, 1973). Ob jemand qualifiziert oder nicht sei, könne nicht über Wissen oder Intelligenz, sondern über Kompetenz definiert werden (vgl. McClelland, 1973, S. 1). Allerdings sind Kompetenzen nur schwer zu erheben. Es besteht weitgehende Einigkeit darüber, dass dies, wenn überhaupt, nur über stellvertretende Facetten von Kompetenzen gelingen kann. Denn Kompetenzen sind Produkte kognitiver Fähigkeiten und Fertigkeiten unter Rückgriff auf mentale Ressourcen, die zur Lösung von Aufgaben benötigt werden (Weinert, 2001, S. 46); eine direkte Erfassung von Kompetenz ist somit unmöglich. Zugleich wird diskutiert, ob eine zu starke Fixierung auf das Kognitive dem Handeln, vor dem sich Kompetenzen beweisen müssen, ausreichend gerecht wird (Wilhelm & Nickolaus, 2013, S. 25). Daher ist eine Spezifizierung der Kompetenzen, auch für Schülerinnen und Schüler sowie für Lehrkräfte, nötig. Zur Kompetenzvermittlung an Lernende benötigen diese Lehrkräfte Lehrkompetenz (vgl. Hattie, 2012). Entscheidende Vermittlungsaspekte liefern die deutschen Bildungsstandards. Diese sind in die Kompetenzbereiche Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung aufgeteilt (KMK Standards, 2005a, 2005b). Im Bildungsplan für Baden-Württemberg werden neben kognitiven oder fachbezogenen auch personale, soziale und methodische Kompetenzen berücksichtigt (z. B. Ministerium für Kultus, Jugend und Sport, 2012).

Zur Spezifizierung von Lehrkompetenz braucht es ebenso einen Rahmen, der von möglichst vielen anerkannt wird. International bildet das Modell der Lehrkompetenz nach Lee S. Shulman (siehe Kapitel 2.1.1, S. 19) einen solchen Rahmen. Zumindest im deutschsprachigen Raum befürworten viele die weitere Konkretisierung des Ansatzes von Shulman durch das Kompetenzmodell von COACTIV (siehe Kapitel 2.1.2, S. 21). Dieses ergänzt Shulmans Ansatz durch drei Aspekte professioneller Kompetenz, die auf die Bedeutung der Persönlichkeit der Lehrkraft fokussieren.

Um die komplexen Kompetenzen, die Lehrkräfte benötigen, messbar zu machen, werden indirekte Erhebungsmethoden notwendig sowie Modelle, die solche indirekten Kompetenzerhebungen erklären. Entsprechend wird hier im Weiteren davon ausgegangen, dass über die Pro-

fessionelle Wahrnehmung (Goodwin, 1994) indirekt Aspekte von Lehrkompetenz erhoben werden können (siehe Kapitel 3.2.1, S. 38). Zu diesem Zweck hätte auch beispielsweise das PID-Modell (Blömeke, Gustafsson & Shavelson, 2015) genutzt werden können. Dieses unterscheidet entweder zwischen **Perception** (Bemerken entscheidender Unterrichtssituationen), **Interpretation** (Deutung der wahrgenommenen Unterrichtssituation) und **Decision making** (Entscheidungen treffen zwischen der Vorhersage einer Schülerinnen- und Schülerreaktion oder dem Vorschlag einer alternativen Vermittlungsstrategie) (Blömeke et al., 2015, S. 7); an anderer Stelle wird entsprechend zwischen **Perceiving**, **Interpreting** und **Decision making** (vgl. Santagata & Yeh, 2015, S. 154) unterschieden (vgl. Abbildung 1).

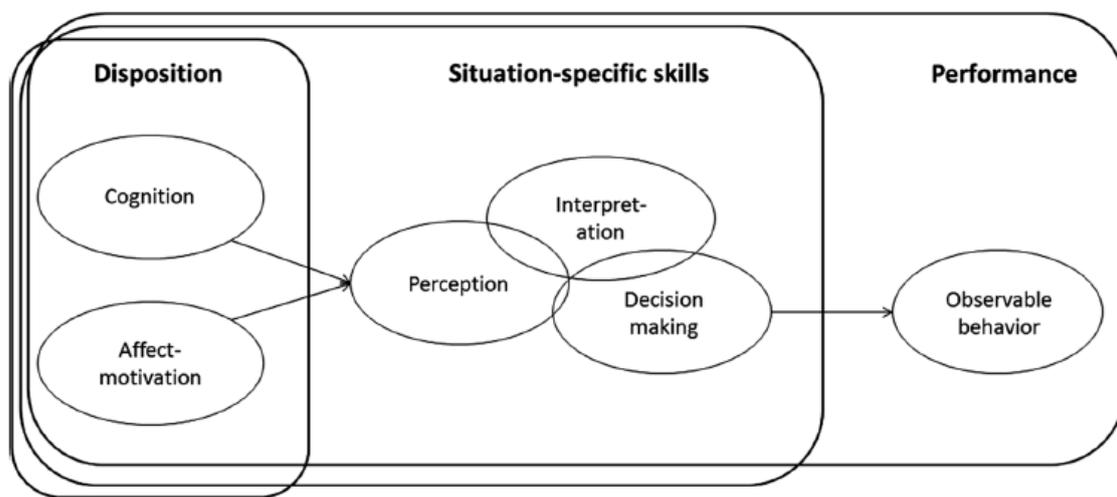


Abbildung 1: PID-Modell nach Blömeke et al. (2015, S. 7).

Wie Abbildung 1 zu entnehmen ist, werden dem PID bei Blömeke et al. (2015, S. 7) die Dispositionen „Kognition“ und „Affektmotivation“ vorgeschaltet und aufbauend auf PID die Performanz als „beobachtbares Verhalten“ bezeichnet. Im Gegensatz dazu wird bei Santagata und Yeh (2015, S. 163) *Wissen und Überzeugungen* sowie *Unterrichtspraxis* als Idee von Blömeke et al. (2015) aufgegriffen und um die Aspekte *Schülerinnen und Schüler*, *Werkzeuge* und *Lehrkräfte-Gemeinschaften* ergänzt. Hierdurch soll größere Nähe zur Unterrichtspraxis erzeugt werden.

Aufgrund der größeren Bekanntheit und wegen der besseren Passung zur Testanlage fiel in dieser Studie die Entscheidung für das Konstrukt Professionelle Unterrichtswahrnehmung.

2.1 Lehrkompetenzen

Über das, was die Tätigkeit einer Lehrkraft im Kern ausmacht, besteht eine Debatte (Baumert & Kunter, 2006; Helsper, 2007). Mindestens zwei Sichtweisen werden unterschieden: Bei der *strukturtheoretischen* Sichtweise wird an die erste Stelle der Profession von Lehrkräften die Vermittlung von Wissen gestellt (Oevermann, 1996, S. 145). Weitere Aufgaben seien Normvermittlung und „implizit therapeutische“ (Oevermann, 1996, S. 146) Tätigkeiten, die notgedrungen zu einem Scheitern in „Handlungssituationen“ (Oevermann, 1996, S. 160) führe. Dieses Scheitern sei der Normalfall (vgl. Oevermann, 1996, S. 160). Im empirisch-quantitativen Paradigma ist die strukturtheoretische Sichtweise nur schwer umzusetzen. Denn wie soll mit der zum Scheitern führenden implizit therapeutischen Dimension des Lehrberufs umgegangen werden und wie könnte diese operationalisiert werden?

Bei der anderen Sichtweise, der *professionellen Handlungskompetenz* (Baumert & Kunter, 2006), wird zunächst vom Gelingen des im Fokus stehenden Unterrichts ausgegangen und dieser in Abhängigkeit von Lehrkompetenzen gestellt (vgl. Hattie, 2009). Dieser Ansatz wird im Folgenden weiter verfolgt, da hierdurch eine empirische Operationalisierung ermöglicht wird.

Professionalisierung wird von Baumert und Kunter (2006) als Beschreibung von Zuständen begriffen. Die Entstehung von Professionalität bei Lehrkräften kann mithilfe verschiedener Modelle erklärt werden (als Überblick: Jong, van Driel & Verloop, 2005b; Justi & van Driel, 2005b). Sprinthall, Reiman und Thies-Sprinthall (1996) beschreiben beispielsweise drei Schritte möglicher Professionalisierung als Prozess: Lehrprofessionalität könne als ein *Handwerk* (Lehrerfahrungs-Modell), als *Lehre* von Novizinnen und Novizen⁴ durch Expertinnen und Experten (Expertenmodell) und als eine *Verbindung* von *institutionalisierter Lehre mit eigenen Lehrerfahrungen* durch die Lehrkraft (interaktives Modell) verstanden werden.

Clarke und Hollingsworth (2002, S. 948) beschreiben hingegen Professionalisierung als Folge von Training, Adaption, persönlichem Fortschritt (Suche nach Veränderung), lokaler Reform (persönlichem Wachstum), systematischer Restrukturierung und Wachstum bzw. Lernen. Sie schlagen vor diesem Hintergrund ein prozessorientiertes und ineinandergreifendes Modell zur Beschreibung der Zunahme der Lehrprofessionalität (Interconnected Model of Teacher Professional Growth - IMTPG) vor (siehe Abbildung 2).

⁴ Anfängerinnen und Anfänger werden im Weiteren mitunter als Novizinnen und Novizen bezeichnet. Für das dahinter stehende Novize-Experten-Konzept vgl. z. B. Freire und Fernandez (2014).

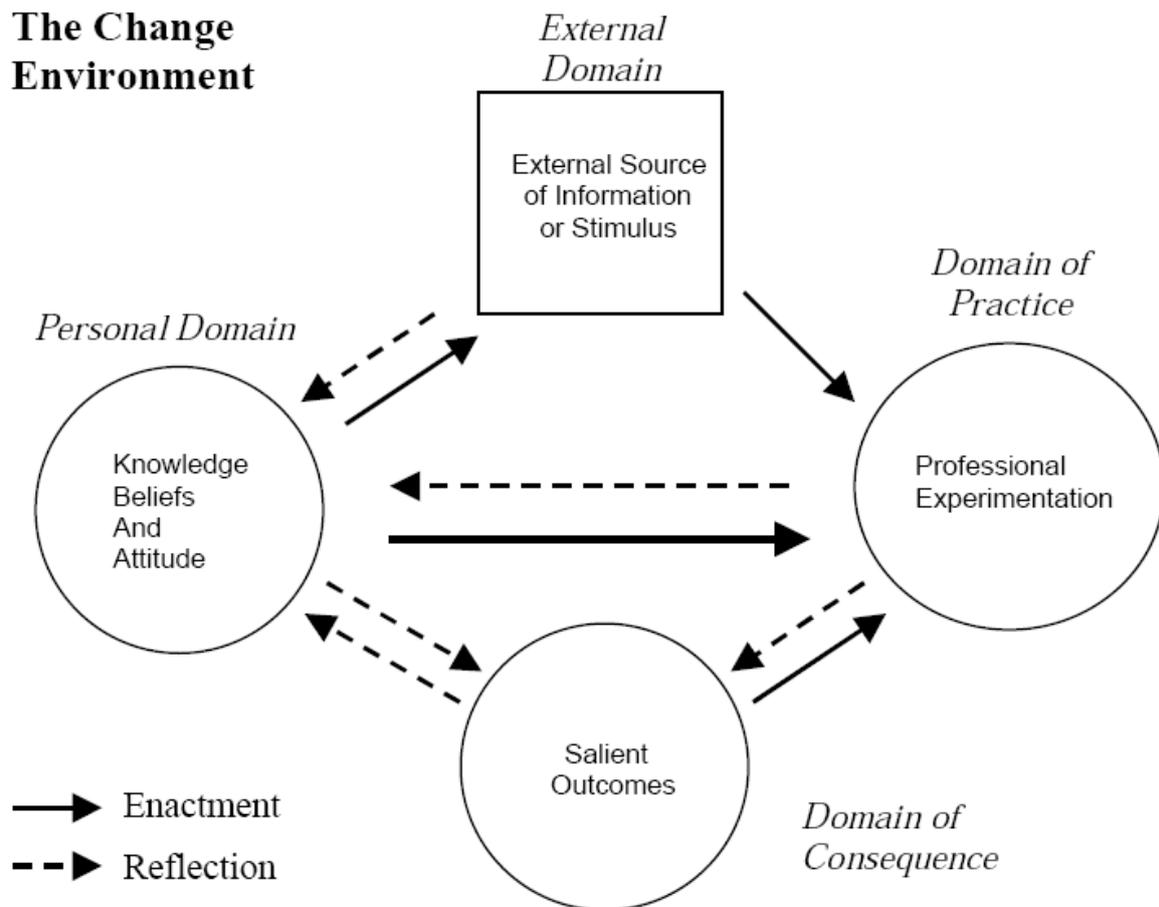


Abbildung 2: Interconnected Model of Teacher Professional Growth – IMTPG (Clarke & Hollingsworth, 2002, S. 948).

Die Fähigkeit zu unterrichten gründet sich sowohl bei Anfängerinnen und Anfängern als auch bei erfahrenen Lehrkräften laut Clarke und Hollingsworth (2002) im Wesentlichen auf vier Domänen sowie deren Auswirkungen: erstens auf eine *externe Domäne*, welche zu Informationen und Anregungen führe; zweitens auf eine *Domäne der Unterrichtspraxis*, welche neue Unterrichtsansätze erprobt; drittens auf eine *Domäne der Konsequenzen*, bei der herausragende Unterrichtserfolge wahrgenommen werden und viertens auf eine *persönliche Domäne*, welche über die wahrgenommene Selbstwirksamkeit zu positiven Wissensüberzeugungen und Haltungen führe.

2.1.1 Kategorien des Professionswissens nach Shulman

Es kann neben dem Prozess der Professionalisierung auch rein deskriptiv erörtert werden, welche Bestandteile Lehrkompetenz beinhaltet. Hierzu ist das meistrezipierte Lehrkompe-

tenzmodell das von Shulman (1986, 1987). Drei Aspekte von Lehrkompetenz stehen im Mittelpunkt der Betrachtung (vgl. Shulman, 1991, S. 159):

- subject matter content knowledge (kurz CK)
- general pedagogical knowledge (kurz PK)
- pedagogical content knowledge (kurz PCK).

CK wird definiert als „Umfang und Organisation von Wissen im Kopf des Lehrers“ (Shulman, 1991, S. 150) und als Fachwissen übersetzt. Der englische und deutsche Begriff gelten weitgehend als deckungsgleich (Bromme, 1995; Gramzow, Riese & Reinhold, 2013).

PK (Shulman, 1987), als allgemeindidaktisches Wissen oder auch als pädagogisch-psychologisches Wissen (PPK; vgl. Voss & Kunter, 2011) bezeichnet, umfasst das generelle Wissen der Lehrkraft darüber, was Schülerinnen und Schüler benötigen, um gut lernen zu können. Übersetzt als allgemeine Didaktik beinhaltet PK Aspekte, die allen Lehrenden, in allen Fachrichtungen und Altersstufen Ähnliches abverlangen, wie etwa die Organisation der Klasse. Shulman (1987, S. 8, Übersetzung B. T.) präzisiert PK als „allgemeine Prinzipien und Strategien [...],⁵ welche über das Fachwissen hinauszugehen scheinen“.

PCK ist ein Wissen, das Verknüpfungen von Fachwissen, pädagogischem und kontextuellem Wissen ermöglicht (vgl. Gess-Newsome, 1999, S. 10). Als „spezielles Amalgam“ (Shulman, 1987, S. 8, Übersetzung B. T.) wird die Gewichtung von PK und CK gegenüber PCK diskutiert. PCK enthält sowohl deklaratives (vgl. Fußnote 2, S. 13) als auch prozedurales⁶ Wissen (Baumert & Kunter, 2006, S. 482) und wird mit fachdidaktischem Wissen übersetzt. PCK schließt auch Kenntnisse darüber mit ein, was das „Erlernen eines bestimmten Themas leicht oder schwer macht“ (Shulman, 1991, S. 151). So umfasst es beispielsweise auch das Wissen, welche Alltagskonzepte Schülerinnen und Schüler in einem bestimmten Bereich haben können und welche „erklärungskräftigsten Methoden“ (Shulman, 1991, S. 151) am hilfreichsten sind, diese zu durchbrechen oder umzuformen. Zu den erklärungskräftigen Methoden zählt auch ein adäquater Umgang mit Modellen, da dieser die Chance bietet, zur Erkenntnisgewinnung in den Naturwissenschaften beizutragen und Präkonzepte umzuformen.

Oft wird nicht beachtet, dass aufgrund unterschiedlicher historischer Entwicklungen durchaus Differenzen im Begriffsverständnis zu PCK bestehen. Fensham (2004, S. 145), aber auch

⁵ Eckige Klammern werden im Folgenden bei wörtlichen Zitaten verwendet, wenn z. B. nur die grammatikalische Form verändert oder Satzzeichen ergänzt werden, um das Zitat an den Fließtext anzupassen.

⁶ Nach Gruber (2008) wird unter prozeduralem Wissen eine höhere Wissensstufe der ACT-Theorie verstanden, die über dem deklarativen (vgl. Fußnote 2, S. 11) Wissen angesiedelt ist und sich mit der Fertigkeit zu wissen, wie etwas funktioniert bzw. zu verstehen ist, beschäftigt.

Hopmann und Riquarts (1995), Bromme (1995) und Gramzow et al. (2013) weisen auf grundsätzliche Unterschiede zwischen Fachdidaktik und PCK hin. Während im angelsächsischen Bereich eine Curriculum-Tradition bestünde, würde in vielen kontinentaleuropäischen Ländern von der Didaktik bzw. von der Bildungstradition ausgegangen. Entsprechend wird auch PCK unterschiedlich verstanden.

Bei Shulman (1987, S. 8, Übersetzung B.T.) findet sich neben CK, PK und PCK zudem das „curricular knowledge“ bzw. Wissen über das Fachcurriculum, das „Wissen über Lernende und ihre charakteristischen Merkmale“, das „Wissen über den Kontext der Bildung, von den Arbeiten in Gruppen oder im Klassenzimmer, die Leitung und Finanzierung der Schulbezirke bis zu den Charakteren der Gemeinschaften und Kulturen reichend“ und das „Wissen über das Beenden, die Zwecke und die Bedeutungen der Ausbildung, und ihre philosophischen und historischen Gründe“. Diese Aspekte werden jedoch häufig nicht näher betrachtet oder als eigene Bereiche, z. B. unter dem Begriff „Organisationswissen“ (Baumert & Kunter, 2011), geführt.

2.1.2 Kompetenzmodell von COACTIV

Auch Baumert und Kunter (2011) gründen ihr Lehrkompetenzmodell auf die Vorarbeiten Shulmans (1986, 1987), mit den erwähnten Einschränkungen in Bezug auf die Gleichsetzung von Begriffen aus Curriculum- und Didaktiktradition. Beim Projekt COACTIV (Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die Entwicklung mathematischer Kompetenz) werden Shulmans Kompetenzbereiche Fachwissen (CK), Pädagogisch-psychologisches Wissen (PPK) und fachdidaktisches Wissen (PCK) um Organisations- und Beratungswissen sowie um Kompetenzfacetten ergänzt. Wichtig erscheint insbesondere die Ergänzung der Aspekte professioneller Lehrkompetenz. Ergänzt werden hier nicht nur das viel beforschte Professionswissen, sondern auch Überzeugungen, Werthaltungen und Ziele, motivationale Orientierungen und Selbstregulation (siehe Abbildung 3).

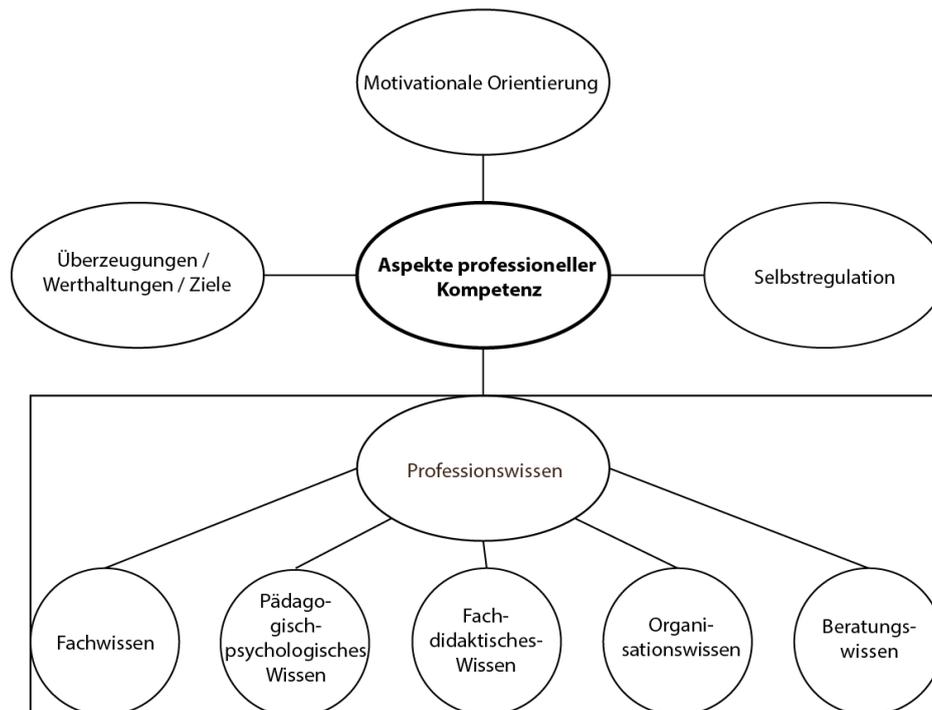


Abbildung 3: COACTIV-Kompetenzmodell nach Baumert und Kunter (2011, S. 32), eigene Darstellung.

2.2 Quantitative Studien zur Lehrkompetenzstruktur

Einige Largescale-Studien⁷ aus dem Bereich der Mathematikdidaktik untersuchen den Einfluss von verschiedenen Variablen auf das Professionswissen von Mathematiklehrkräften. Hier werden zwei zentrale Studien vorgestellt: zum einen COACTIV, zum anderen MT21. Diese beiden Studien dienen aufgrund ihrer psychometrischen und theoretischen Fundiertheit zum einen als Vorbild, zum anderen zur Kontrastierung. Die in dieser Studie entwickelten Tests unterscheiden sich von den beiden Largescale-Erhebungen durch ihre Situiertheit und ihre Nähe zur Unterrichtsrealität. Die These lautet, dass eines der zentralen Ergebnisse beider Largescale-Erhebungen, nämlich hohe latente Korrelationen (Kovarianzen) zwischen dem Fachwissen CK und dem fachdidaktischen Wissen PCK, auch mit der Art des Testens einhergeht.

2.2.1 COACTIV

Das Forschungsprojekt COACTIV (Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die Entwicklung mathematischer Kompetenz) verbindet Erhebung-

⁷ Mit Largescale-Studien sind im Folgenden Studien mit großen Stichproben gemeint.

gen zur PISA-Studie (Schülerebene) mit Lehrkompetenzmessungen. Besonderes Interesse liegt auf den Aspekten adäquate Klassenführung, kognitive Aktivierung und Motivation. Ergänzend zu PISA 2003/2004 wurden deutsche Mathematiklehrkräfte aller Schulformen (ohne Sonderschulen) der neunten und zehnten Jahrgangsstufe in die Untersuchung einbezogen. Auf eine Befragung von $N = 351$ Lehrkräften im Jahr 2003 folgte eine Befragung im Jahr 2004 mit $N = 229$ Mathematiklehrkräften. Die Tests bestanden aus offenen Antwortmöglichkeiten, um ein Erraten der richtigen Antworten zu verhindern (Brunner et al., 2006, S. 64). Referenz für die Raterinnen oder Rater waren entweder Ergebnisse von anderen Studien oder Antworten von Expertinnen und Experten (Brunner et al., 2006, S. 66). Die Aspekte professioneller Kompetenz, nämlich Überzeugungen und „Beliefs“, motivationale Orientierungen und Selbstregulation wurden durch Selbstberichte erhoben (Brunner et al., 2006, S. 67). Der tatsächliche Unterricht wurde über Lernenden- und Lehrkräftebefragungen sowie Klassenarbeitsaufgabenanalysen erfasst (Brunner et al., 2006, S. 67)

Bei COACTIV wurden Testaufgaben entwickelt, die nur in geringem Umfang Unterricht situativ wiedergeben. Psychometrisch entstand so ein Test, der den Erfordernissen der Reliabilität (auch der Interraterreliabilität) entspricht (Brunner et al., 2006, S. 66). Die Aufgaben zum fachdidaktischen Wissen sind beispielsweise wie folgt konstruiert:

„Eine Schülerin berechnet für die Gleichung $(x - 3)(x - 4) = 2$

die Lösungen $x = 5$ oder $x = 6$. Was hat diese Schülerin vermutlich gerechnet?“ (Krauss et al., 2011, S. 140)

Aufgaben wie diese lassen sich auch mit mathematischem Fachwissen lösen. Entsprechend hoch sind die latenten Korrelationen zwischen Fachwissen und fachdidaktischem Wissen. Diese werden für alle Lehrkräfte mit .79 (Krauss et al., 2011, S. 148), für Lehrkräfte, die nicht aus dem Gymnasium stammen, mit .61, für Lehrkräfte aus dem Gymnasium sogar mit .96 (Krauss et al., 2011, S. 149) berichtet. Offen bleibt, ob sich die hohe Korrelation aus Fachwissen und Fachdidaktik aus der nahen Operationalisierung speist oder die beiden Konstrukte tatsächlich so stark zusammenhängen.

Auf Grundlage des COACTIV-Tests ergeben sich folgende Aussagen: An Gymnasien aktiviere Unterricht am stärksten kognitiv. Dies geschehe an Hauptschulen am wenigsten. Hauptschullehrkräfte führten die Schülerinnen und Schüler am engsten. Hingegen werde am Gymnasium am stärksten individualisiert und differenziert (vgl. Brunner et al., 2006, S. 72). Außerdem führe Unterrichtserfahrung nicht notwendig zu besserem PCK und CK bei Lehrkräften (Brunner et al., 2006, S. 75). Die Strukturierung des eigenen Lernprozesses und tatsächli-

che Gelegenheiten zum Lernen seien für Lehramtsstudierende entscheidend für die Entwicklung des Professionswissens (Brunner et al., 2006, S. 76).

2.2.2 MT21

Die Studie *Mathematics Teaching in the 21st Century (MT21)* vergleicht international zwischen den Ausbildungen von Mathematiklehrerinnen und Mathematiklehrern der Länder Bulgarien, Taiwan, Deutschland, Südkorea, Mexiko und den Vereinigten Staaten von Amerika (Schmidt et al., 2008, S. 483), ergänzt um national vertiefte Fragestellungen. $N = 849$ Lehramtsstudierende und Vorbereitungsdienstleistende der Mathematik wurden in Deutschland querschnittlich befragt (Blömeke, Felbrich & Müller, 2008, S. 34).

In den in Deutschland durchgeführten Tests schneiden künftige Lehrkräfte für Gymnasien und Gesamtschulen deutlich besser ab als künftige Lehrkräfte der Grund-, Haupt- und Realschulen. Dies gilt insbesondere für den fachwissenschaftlichen Bereich, aber auch, wenn gleich schwächer, für den fachdidaktischen (Blömeke, Kaiser, Schwarz et al., 2008, 92 f.). Dem Geschlecht der künftigen Lehrenden scheint hierbei kein signifikanter Erklärungswert zuzukommen (Blömeke, Lehmann et al., 2008, S. 121), jedoch der Dauer des Studiums und des Referendariats (Blömeke, Kaiser, Schwarz et al., 2008, S. 103). Entscheidend für das Abschneiden sind weiterhin die Abiturnote und die Wahl des Leistungskurses; beides hat starke Auswirkungen auf das Vermögen, die Tests gut zu lösen (Blömeke, Lehmann et al., 2008, S. 121).

Da zumeist geschlossene Items verwendet wurden, bestehen Zweifel hinsichtlich der Eigenständigkeit der Gedankengänge der künftigen Lehrkräfte. Es wird eine hohe „(messfehlerbereinigte) latente Korrelation zwischen Mathematik und Mathematikdidaktik“ von .81 (Blömeke, Kaiser & Lehmann, 2008, S. 69) berichtet.

Wegen Zweifeln an den hohen Korrelationen wurden zusätzlich sogenannte Bridging-Items entwickelt. Diese sind Fragen, welche stärker auf die Messung der Handlungskompetenz der künftigen Lehrenden abzielen und fachwissenschaftliche und fachdidaktische Elemente in sich vereinen (Marx & Rinkens, 2008, S. 443). Abweichend von den in der Hauptuntersuchung dargestellten engen Zusammenhängen zwischen fachwissenschaftlichem und fachdidaktischem Wissen zeigen sich bei $N = 112$ Probandinnen und Probanden (Grund-, Haupt- und Realschullehramtsstudierende) gänzlich andere Ergebnisse. „Durch die Bündelung des fachlichen und fachdidaktischen Wissens innerhalb eines inhaltlichen Kontextes“ wurde „eine deut-

lichere Trennung dieser beiden Wissensbereiche“ (Marx & Rinkens, 2008, S. 450) gemessen. Marx und Rinkens (2008, S. 450) plädieren deshalb für „inhaltsnahe[] Items“.

2.3 Vignettentests: ökologisch generalisierbare Kompetenzmessung

Sowohl die Studienergebnisse aus COACTIV als auch aus MT21 werfen Fragen nach der ökologischen Generalisierbarkeit (Messick, 1989) auf. Bei dieser geht es um den Anspruch auf Realitätsnähe und auf die Übertragbarkeit des Testens in die Praxis (vgl. Kromrey & Renfrow, 1991; McClelland, 1973). Messick (1989, S. 46) steht Tests kritisch gegenüber, die für sich beanspruchen, ökologisch Generalisierbares zu messen. Er argumentiert, dass bei solchen Tests nicht bedachte andere Konstrukte konfundieren könnten (vgl. MNAR, Kapitel 5.3, S. 72) und somit die Übertragbarkeit zu herkömmlichen Tests nicht unbedingt gegeben sein müsse.

2.3.1 Der Vignettentest als Forschungsinstrument

Als spezielles Format eines Forschungsinstruments steht der Vignettentest zur Verfügung. Ursprünglich wurden mit Vignetten (französisch *vigne*: „Weinrebe“) Schilder mit Rebsorten- und Eigenschaftsbeschreibungen entlang der Wirtschaftswege im Weinberg bezeichnet (vgl. Wikipedia, 2016). Aus solchen Beschreibungen von Eigenschaften und Zuständen ist das Forschungsformat Vignette entstanden, bei dem ein als entscheidend erachteter Ausschnitt von Realität stellvertretend präsentiert wird. Auf diesen müssen die Probandinnen und Probanden in einer bestimmten Weise reagieren.

Vignettentests werden in verschiedenen Disziplinen verwendet. Indem eine realitätsnahe Situation präsentiert wird, soll auf einem Kontinuum zwischen Wissen und Kompetenz das Konstrukt Kompetenz stärker fokussiert werden. So werden z. B. beim Staatsexamen am Ende des Medizinstudiums bestimmte Krankheitsbilder von Patientinnen und Patienten fallbasiert präsentiert. Die Annahme, dass durch die korrekte bzw. falsche Reaktion auf die im Test dargestellte (Krankheits-)Situation Rückschlüsse auf das Handeln möglich sind, unterscheidet den Vignettentest von der reinen Faktenabfrage, erlaubt aber dennoch quantifizierbare Aussagen.

Wenn hier Vignettentests thematisiert werden, so ist eine Sonderform des Beobachtens von Unterricht gemeint. Direkte Unterrichtsbeobachtung lässt zwar ein direktes Erfassen der Lehrkompetenz zu und kann deshalb verhältnismäßig valide messen, ist jedoch forschungsökonomisch sehr aufwendig. Schließlich müssen Unterrichtsstunden oder Videos von Unter-

richtstunden von Expertinnen und Experten bzw. von Raterinnen und Ratern bewertet werden. Außerdem ergibt sich häufig das Problem, dass die beobachtenden Personen dem komplexen Gebilde Unterricht die nur schwer zu trennenden Beobachtungsgesichtspunkte nicht klar zuweisen können.

Hier nimmt der Vignettentest eine spezifische Stellung ein. Probandinnen und Probanden werden vor reale bzw. in der Schulrealität denkbare Situationen und Probleme gestellt. Nicht das eigene Verhalten wird reflektiert, sondern die Aktivitäten der handelnden Lehrperson werden stellvertretend analysiert. Mitunter wird versucht, die reale Schulsituation zu simulieren, indem zur Bearbeitung der Fragen zeitliche Beschränkungen eingeführt werden. Vignettentests sind somit kontextabhängige Beschreibungen von Unterrichtssituationen (Stecher et al., 2006, S. 102), zu denen Probandinnen und Probanden in Form von geschlossenen oder offenen Items schriftlich oder mündlich Stellung nehmen und Lösungsvorschläge entwickeln (vgl. Rehm & Bölsterli, 2014). Vignetten können über verschiedene Medienformen präsentiert werden, z. B. Grafiken oder Comics. Gemeinhin werden Vignettentests jedoch in Text- und Videovignetten eingeteilt. Textvignetten präsentieren die Unterrichtssituationen als Text und gelten deswegen als effizient (vgl. Jeffries & Maeder, 2011, S. 162). Das Maß an Nähe zur Unterrichtsrealität ist durch die authentische Repräsentation bei Videovignetten jedoch höher (Seidel, Blomberg & Stürmer, 2010, S. 305). Gleichzeitig wird diskutiert, ob die Fokussierung auf *entscheidende* Unterrichtssituationen bei Videovignetten schwieriger wird, insbesondere für Novizinnen und Novizen⁸ (Syring et al., 2015). Steffensky, Gold, Holdynski und Möller (2015, S. 353) unterstützen dies und führen die Divergenzen zwischen Novizinnen und Novizen sowie Expertinnen und Experten auf Unterschiede im Umgang mit kognitivem Schemata und Skript zurück. Entsprechend berichten Syring et al. (2015, S. 676) bei einer Videovignettengruppe von signifikant höherer kognitiver Belastung als bei einer Textvignettengruppe.

Zur Fokussierung und damit zur Reduzierung der kognitiven Belastung werden im Gegensatz zu früheren Vignettentests heute eher nachgedrehte bzw. artifizielle Videovignetten präsentiert anstelle von videografiertem, tatsächlichem Unterricht (Oser, Heinzer & Salzmann, 2010). Die Befunde hierzu sind jedoch widersprüchlich.

Es ist davon auszugehen, dass Textvignetten etwas anderes messen als Videovignetten (Syring et al., 2015). Wie genau sich in Abhängigkeit von der Herstellung der Videovignetten das

⁸ Vgl. Fußnote 4, S. 18.

Messergebnis unterscheidet ist jedoch nicht geklärt und vermutlich abhängig vom jeweiligen Vignettentest.

Gemeinsam ist den Vignettentests, dass bei der Auswertung als Referenzwert ein Expertinnen- und Expertenurteil genutzt wird, das mit dem Ergebnis der Probandin oder des Probanden verglichen wird. Als lehrkompetent gilt, wer viele Übereinstimmungen mit dem Expertinnen- und Expertenurteil aufweisen kann.

2.3.2 Vignettentests im Vergleich

Um einen Einblick in das Forschungsinstrument Vignettentest zu bekommen, werden hier einige vorgestellt. Da sich der Vignettentest in den letzten Jahren in der Lehrkompetenzforschung stark verbreitet hat, ist eine vollständige Aufzählung nicht möglich. Um die Entwicklung zu zeigen, werden hier zunächst ältere Vignettentests vorgestellt und am Ende eine Tendenz aufgezeigt. Insbesondere wird das Konstrukt, das mit dem Vignettentest erhoben werden soll, dargestellt.⁹

Oser et al. (2010) entwickeln ihren Videovignettentest aus einer Kritik an vormals üblichen Lehrkompetenztests. Diese würden auf die zentralen Aspekte des Lehrens, die „Situativität, Authentizität, Komplexität und (...) Kontextgebundenheit unterrichtlichen Handelns“ (Oser et al., 2010, S. 5) nicht eingehen. Um diesem Anspruch gerecht zu werden, können über Beobachtung und Beurteilung durch die Probandinnen und Probanden Rückschlüsse auf deren Kompetenzen gezogen werden (Oser et al., 2010, S. 9). Oser et al. (2010) nennen das mit Vignettentests gemessene Konstrukt „*advokatorischer Ansatz*“. Dieser ermögliche Aussagen über Kompetenzprofile.

Der Videovignettentest von Oser et al. (2010) enthält Videos tatsächlicher Unterrichtssituationen. Oser et al. (2010) lassen die Probandinnen und Probanden nach dem Betrachten der Videos Items auf einer sechsstufigen Likertskala bewerten. Die Antworten der Probandinnen und Probanden werden mit Expertinnen- und Expertenaussagen verglichen.

Baer et al. (2007) kombinieren Video- und Textvignetten. Die gemessenen Konstrukte nennen sie „*unterrichtliche Handlungskompetenz*“ (Baer et al., 2007, S. 29), Situationen des Vorbereitens von Unterricht und „speziell[e] diagnostische Kompetenz[en]“ (Baer et al., 2007, S. 28). Im Gegensatz zu Oser et al. (2010) handelt es sich hier um ein offenes Aufgabenfor-

⁹ Die mit Vignetten erhobenen Konstrukte werden kursiv geschrieben.

mat, bei dem Probandinnen- und Probandenantworten durch Raterinnen und Rater mithilfe eines Manuals analysiert wurden. Hierbei wurden $N = 46$ bzw. $N = 32$ Personen befragt. Bei den Textvignetten zeigen sich hochsignifikante Zunahmen des Mittelwerts im Vergleich zwischen den didaktischen Fähigkeiten im ersten und dritten Semester des Studiums (Baer et al., 2007, S. 35) und signifikante Zunahmen des Mittelwerts bei den diagnostischen Fähigkeiten. Dagegen wurde weder in der Klassenführung noch in der Sachkompetenz eine signifikante Zunahme gemessen (Baer et al., 2007, S. 36). Bei den Erhebungen durch Videovignetten zeigen sich hochsignifikante Zunahmen des Mittelwerts im didaktischen Verständnis und signifikante Zunahmen der Klassenführungskompetenz (Baer et al., 2007, S. 38). Allerdings zeigen sich keine signifikanten Veränderungen im diagnostischen Bereich und in der Sachkompetenz.

Im Forschungsprojekt *Observe* (vgl. Seidel et al., 2010) wurden Videovignetten validiert und standardisiert. Das untersuchte Konstrukt ist die *Professionelle Wahrnehmung*, die in *Noticing* und *Reasoning* aufgeteilt wird (Seidel et al., 2010, S. 298). Die zwölf Videoclips entstammen realem Unterricht der Fächer Physik, Mathematik, Geschichte, Französisch und Englisch (Seidel et al., 2010, S. 299). Es werden vier- oder fünfstufige Skalen benutzt, bei denen ein Urteil von Expertinnen und Experten (Cohens Kappa = 0.79) den Vergleichswert bildet. Zunächst wurden $N = 40$ (Seidel et al., 2010), später $N = 387$ (Jahn, Prenzel, Stürmer & Seidel, 2011), schließlich $N = 1029$ (Jahn, Stürmer, Seidel & Prenzel, 2014) Lehramtsstudierende befragt. Zentrale Ergebnisse sind insbesondere die Standardisierung, Hinweise auf die (Inhalts- und Konstrukt-)Validität und die IRT-basierte Passung des theoretischen Modells zu den Daten (Jahn et al., 2014).

Riese und Reinhold (2010, S. 169) stellen „ein Technologiedefizit bei der angemessenen Erfassung von Kompetenzen“ fest. Dies nehmen sie zum Anlass, neue Forschungsinstrumente zu entwickeln, darunter auch einen Textvignettest. Das untersuchte Konstrukt nennen sie *professionelle Handlungskompetenz* bzw. „*Beurteilung, Analyse und Reflexion von Lernprozessen*“ (Riese & Reinhold, 2010, S. 175). Die Textvignetten bestehen aus Unterrichtsbeschreibungen und Dialogaufzeichnungen mit offenen Items, die von Raterinnen und Ratern klassifiziert wurden. $N = 301$ Lehramtsstudierende ($n = 120$ Haupt- bzw. Realschullehramt, $n = 176$ Gymnasiallehramt, $n = 5$ ohne Angabe) (Riese & Reinhold, 2010, S. 176) wurden befragt. Zentrales Ergebnis ist insbesondere die Bestätigung kriterialer Validität. So zeigen sich „signifikant höhere Testleistungen“ bei „Gruppen mit höherer Expertise“ (Riese & Reinhold, 2010, S. 176).

Schmelzing, Wüsten, Sandmann und Neuhaus (2010) haben neben einem Test zum deklarativen fachdidaktischen Wissen ebenfalls einen Vignettentest mit fiktionalen Unterrichtsvideos entwickelt und eingesetzt. Das gemessene Konstrukt nennen sie *Reflexionsfähigkeit*, operationalisiert als Skalen zu *Aufmerksamkeitsfokus* (Cronbachs $\alpha = .70$; ICC = .75 bei offenem Itemformat und Rating) (Schmelzing et al., 2010, S. 195), „*Bewertung*“ (geschlossenes Antwortformat mit der Abfrage, ob der Unterricht gelungen oder nicht gelungen war, ohne Reliabilitätswerte berichtet) (Schmelzing et al., 2010, S. 197) und „*Analysefähigkeiten*“ (Cronbachs Alpha = .73; ICC = .63 bei offenem Itemformat) (Schmelzing et al., 2010, S. 197). Es wurden $N = 93$ Probandinnen und Probanden ($n = 49$ erfahrene Biologielehrkräfte, $n = 22$ Referendarinnen und Referendare und $n = 22$ Biologiestudierende) befragt. Erwartungskonform werden signifikante Unterschiede entlang der Expertise gemessen, mit Ausnahme der Skalen *Aufmerksamkeitsfokus* und *Analyse*, bei denen die Referendarinnen und Referendare einen höheren Wert aufweisen als Lehrkräfte (Schmelzing et al., 2010, S. 201). Grundsätzlich werden im Vergleich höhere Werte bei Lehrkräften als bei Studierenden festgestellt.

Das Projekt „Naturwissenschaften unterrichten können“ (NUK) untersucht mit einem Textvignettentest, bestehend aus acht Vignetten, das Können von Lehramtsstudierenden der Naturwissenschaften Physik, Chemie und Biologie in der Schweiz und in Deutschland. Zumindest dem Namen nach startete NUK mit dem Anspruch, *Handlungskompetenz* zu untersuchen. Dieses Ziel wurde durch das Konstrukt ersetzt, *authentisch abgebildete Kompetenzen vor dem Hintergrund „größerer Unterrichtszusammenhänge“* (Brovelli, Bölsterli, Rehm & Wilhelm, 2013, S. 307) abzubilden. Um möglichst nahe am tatsächlichen Unterricht zu sein und die Antwort so wenig wie möglich zu beeinflussen, wurde auf die Konkretisierung der Fragen verzichtet. Stattdessen wurde nur um stichwortartige „Rückmeldung“ gebeten (Brovelli et al., 2013, S. 315). Es wurden $N = 344$ Probandinnen und Probanden unterschiedlicher Studiedauer befragt und die Antworten durch Raterinnen und Rater mit guten Interraterreliabilitäten bepunktet (nach den Kompetenzbereichen CK, PCK und PK). Es konnte ein Kompetenzzuwachs über die Semester in allen Unterskalen in CK, PCK und PK gezeigt werden (Brovelli et al., 2013, S. 321).

Kersting et al. (2012) präsentierten 13 Videovignetten zum Mathematikunterricht, die mithilfe von offenen Items bewertet werden. Sie entwickelten den Vignettentest auf der Grundlage des Konstrukts „*im Klassenzimmer brauchbares und nutzbares Wissen*“, mit dem sie sich ausdrücklich von Hill, Ball, Blunk, Goffney und Rowan (2007) und dem Konstrukt „*Mathematical Knowledge for Teaching*“ (MKT) absetzen. Es wurden $N = 237$ Mathematiklehrkräfte über das Internet befragt. Die Antworten wurden hinsichtlich mathematischem Inhalt, dem

Denken von Schülerinnen und Schülern, Überlegungen zur Verbesserung des Unterrichts und Überlegungen zur Verbesserung der Instruktion geratet. Außerdem wurde bei einem großen Teil der Probandinnen und Probanden ($N = 223$) der MKT-Test eingesetzt (Hill et al., 2007). Es zeigten sich mittlere Korrelationen zwischen den Ratingergebnissen und dem MKT-Test ($r(223) = .608, p < .01$). Die Überlegungen zur Verbesserung des Unterrichts können als Prädiktor für den Erfolg von Schülerinnen und Schülern herangezogen werden ($\beta = .521, p < .05$), der MKT-Test jedoch nicht ($r = -.059, p = .82$) (Kersting et al., 2012, S. 572). Lehrkräfte mit guten Werten bei der Analyse von Videovignetten hinsichtlich des mathematischen Inhalts (MC: $\beta = .624, p < .01$) zeigten in ihrem Unterricht bessere Instruktionsqualität (Kersting et al., 2012, S. 583).

Zusammenfassend ergibt sich folgendes Bild: Ursprünglich wird das mit Vignettentests gemessene Konstrukt *Handlungskompetenz* genannt. Heute besteht weitgehende Einigkeit darüber, dass Kompetenz im unterrichtlichen Handeln mit Vignettentests nicht erhoben werden kann. Stattdessen werden Zusammenhänge von Vignettentests und Unterrichtshandlung berichtet (Kersting et al., 2012). Unter Vorbehalt lässt sich eine Tendenz zur Deskription ausmachen (das Konstrukt wird beschrieben über durch Darbietung von Unterricht angestoßene Prozesse). Diese Prozesse können als Wahrnehmung und Schlussfolgerung (vgl. Kapitel 3.2, S. 38) zusammengefasst werden.

3 Professionelle Unterrichtswahrnehmung des Modelleinsatzes im Unterricht

Wie kann Schülerinnen und Schülern Modellkompetenz vermittelt werden? Mit dieser Frage verbunden ist: Wie kann Vermittlung von Modellkompetenz als Teil der Lehrkompetenz operationalisiert werden? Und wie kann daraus ein Test entstehen, der nahe an der Schulrealität misst? Um diese Fragen zu beantworten, muss zunächst geklärt werden, was Modelle im Biologie- und Chemieunterricht sind und welche Anforderungen an die Modellkompetenz von Schülerinnen und Schülern daraus erwachsen (vgl. Kapitel 3.1, S. 31). Daran schließt sich die Erörterung des Konstrukts *Professionelle Unterrichtswahrnehmung* (vgl. Kapitel 3.2, S. 38 und 3.2.1, S. 38) an. Dieses Konstrukt bietet die Möglichkeit, das Ergebnis des Vignettentests als Aspekt von Modellkompetenz zu beschreiben. Hierauf folgt die Beschreibung der Bestandteile der *Professionellen Unterrichtswahrnehmung* als *Noticing* (selektive Aufmerksamkeit; Kapitel 3.2.3, S. 42) und *Reasoning* (wissensbasierte Schlussfolgerung; Kapitel 3.2.2, S. 40) sowie deren jeweilige Operationalisierung.

3.1 Modelle und Modellkompetenz im Chemie- und Biologieunterricht

Der Mensch erscheint Stachowiak als das „modellbildende Wesen (...). Alles was ihm neu- und fremdartig erscheint, sucht er sich im Medium der Modellbildung anschauend, beobachtend, interpretierend, vergewissernd anzueignen (...). Sein Lernen ist ein Lernen *an* und *mit* Modellen, und sein Handeln wesentlich ein Handeln *nach* Modellen“ (Stachowiak, 1980, S. 9). Doch wie kann Modellkompetenz bei der Fülle verschiedener Modelle vermittelt werden?

In der Chemie und der Biologie gehört das Modellieren wie in allen Naturwissenschaften zu den zentralen Denk- und Arbeitsweisen. Die Modelle im Forschungsgebiet der Chemie erstrecken sich von der Beschreibung großtechnischer Anlagen und Phänomene bis hin zur Subteilchenebene. Ähnliche Modelle sind in der Biologie zu finden, etwa Modelle zum Bioreaktor, Phänotyp oder der DNA. Diese müssen jedoch um Modelle zu komplexen Systemen wie z. B. des Ökosystems oder Immunsystems ergänzt werden. Wegen der größeren „Mannigfaltigkeit (...) der belebten Natur“ (Peters, 1998, S. 102) ist auch von einer größeren Unterschiedlichkeit von Modellen in der Biologie auszugehen.

Aus dieser größeren Bandbreite des Modelleinsatzes in biologischen Zusammenhängen folgt eine geringere Präzision, wenn es darum geht zu definieren, was allen Modellen gemein ist. Zwar besteht in der Chemiedidaktik eine größere Präzision bei der Modelldefinition. Dafür wird in der Biologiedidaktik Präzision über eine größere Bandbreite der Definition von Modelltypen erzeugt.

Daher werden im Folgenden zunächst Definitionen des Modellbegriffs in der Biologie- und in der Chemiedidaktik vorgestellt. Anschließend werden die Modelltypen voneinander abgegrenzt und Modellkompetenz definiert. Alle Modellerläuterungen stehen im Kontext der fachdidaktischen Umsetzung im Unterricht, sind also keine wissenschaftstheoretischen Erläuterungen. Quellen dieser Betrachtung sind Artikel, die über einen systematischen Reviewprozess¹⁰ ausgewählt worden sind.

Unter dem Wort Modell (lat. „modulus“ – Maßstab) werden in der Biologie zum einen „vereinfachte ideelle oder materielle Abbildungen der Wirklichkeit als Ganzes, eines Ausschnittes oder bestimmter Zusammenhänge der Wirklichkeit, die der Veranschaulichung wesentlicher Struktur- oder Funktionsmerkmale originaler Objekte oder Vorgänge dienen“ (Meyer, 1990, S. 4) verstanden. Zum anderen wird der Prozess der *Modellentstehung* in den Vordergrund gerückt. Hier werden die Konstruiertheit, das Vorhersagende und der konzeptuelle Charakter betont (Valanides & Angeli, 2006, S. 88). Beide Definitionen haben ihre Berechtigung. Letztere kann für die Fragestellung dieser Arbeit jedoch als wichtiger erachtet werden, weil sie als Teil des Nature of Science¹¹ anschließt an das bei Schülerinnen und Schülern oft mangelhafte Wissen über die Natur der Modelle (vgl. Krell, Reinisch & Krüger, 2014) und somit für den Schulalltag hilfreicher erscheint.

Crawford und Cullin (2004) schließen an die zweite Definition an und fassen folgende zentrale Modellaspekte zusammen: Modelle sollten als Hilfsmittel zum Verständnis von solchen Phänomenen verstanden werden, deren Richtigkeit überprüfbar ist, die nicht statisch, sondern den jeweiligen Zwecken und Wissensständen unterworfen sind (Crawford & Cullin, 2004, S. 1382). Mit diesem Modellverständnis sollten Schülerinnen und Schüler kompetent werden, Analyse und Synthese naturwissenschaftlicher Theorien zu erkennen (Valanides & Angeli, 2006, S. 88).

¹⁰ Mit dem systematischen Literaturreview ist die systematische, also die begründete Auswahl von passenden 57 Artikeln sowie deren systematische Auswertung gemeint (vgl. Kapitel 3.2.3, S. 42).

¹¹ Mit Nature of Science ist die Natur der Naturwissenschaft (vgl. Kapitel 3.2.3 F, S. 52) gemeint.

Es gibt viele Möglichkeiten der Klassifikation von Modellen im Biologieunterricht. Eine einheitliche Abgrenzung blieb jedoch bisher aus. Litsche (1985, S. 262) weist bspw. auf den Unterschied zwischen „theoretischen“ und „technischen“ Modellen hin. *Theoretische Modelle* existierten „nur in Gedanken, die mittels sprachlicher oder anderer Zeichen ausgedrückt werden“ (Litsche, 1985, S. 262). Deren „nicht-ikonische“ (Litsche, 1985, S. 263) Variante bildet eine Schnittmenge mit Modellen der Chemie und Physik. Eine Abgrenzung zu Analogien bleibt hier ebenso aus wie zum Gedankenmodell, bei dem es sich um „sprachlich formulierte Modellvorstellungen, die auf Beobachtungen oder Experimente zurückgehen“ (Meyer, 1990, S. 5) handelt. Bei *technischen Modellen* handle es sich um anfassbare bzw. „materielle“ Modelle, zu denen das ikonische, häufig auch als Realmodell bezeichnete „Blütenmodell“ (Litsche, 1985, S. 262) gehöre. Bleibt man beim Blütenmodell, so wird die Zuordnungsproblematik deutlich. Nach Schneider (1981, S. 259) könnte man das Modell der Wiesensalbeiblüte (*Salvia pratensis*) wegen seiner Plastizität als *anatomisches Modell* bezeichnen. Stellt man den Klappmechanismus in den Mittelpunkt, könnte das Modell sowohl als *Funktionsmodell* bezeichnet werden als auch als *Realnachbildung*. Ein Funktionsmodell wäre es, wenn man „morphologisch stärker abstrahiert, hinsichtlich einer bestimmten Funktion“ (Schneider, 1981, S. 259). Eine Realnachbildung wäre dieses Modell, falls man den Aspekt „naturgetreu“ in den Fokus nimmt, der sich meist jedoch nicht auf die „Materialbeschaffenheit“ (Schneider, 1981, S. 259) bezieht. Diese Gegenüberstellung zeigt die Grenzen der Abgrenzbarkeit und zugleich die Unschärfe der Begriffe auf.

Ergänzend zu den theoretischen Modellen gibt es die Sonderform der *Simulation*. Simulationen zählen neben Zeichnungen, Analogien und konkreten Modellen zu den am häufigsten verwendeten Modellen im Unterricht (Justi & van Driel, 2005a, S. 200) und zu den wirkungsvollen Unterrichtselementen (Eilks, Witteck & Pietzner, 2009b, S. 53) des naturwissenschaftlichen Unterrichts. Simulationen umfassen eine große Bandbreite von Denkmodellen, mathematischen Formeln, Computerprogrammen (vgl. Angeli, Valanides & Papageorgiou, 2011), Simulationsspielen (z. B. zum Raubtier-Beute-Verhältnis) bis hin zu Funktionstheatern (z. B. Schülerinnen und Schüler spielen die elektrische Reizweiterleitung von Nervenzellen nach). Insbesondere bei den beiden letztgenannten Beispielen wird der Unterschied zu den theoretischen Modellen deutlich: Während bei theoretischen Modellen die adäquate Vermittlung von Theorie im Vordergrund steht, ist es bei diesen szenischen Modellen vielmehr die methodisch bewusst eingesetzte Emotionalisierung, die zu einem tieferliegenden Verständnis des Themas und einer adäquaten Modellvorstellung z. B. von Ursache-Wirkungs-Zusammenhängen füh-

ren soll (vgl. Crawford & Cullin, 2004, S. 1382). Tabelle 1 gibt einen Überblick über Modelltypen, Definitionen und Beispiele im Biologieunterricht.

Tabelle 1: Überblick über Modelltypen im Biologieunterricht

Modelltyp	Definition	Beispiel
theoretisches Modell bzw. Gedankenmodell	<i>Theoretische Modelle</i> existieren „nur in Gedanken, die mittels sprachlicher oder anderer Zeichen ausgedrückt werden“ (Litsche, 1985, S. 262), falls nicht-ikonisch: Überschneidung mit Chemie und Physik. <i>Gedankenmodelle</i> sind „sprachlich formulierte Modellvorstellungen, die auf Beobachtungen oder Experimente zurückgehen“ (Meyer, 1990, S. 5)	Sprachliche Fotosynthesereaktionsgleichung: Kohlendioxyd + Wasser + Licht → Traubenzucker + Wasser + Sauerstoff
technische Modelle	Anfassbare, ikonische Modelle, mit dem Ziel, technisch erklärbare Vorgänge nachzubilden	3D-Nachbildung der Herzklappen. Ziel: Vermittlung der Funktionsweise
anatomisches Modell	Ikonisches Modell, bei dem die Plastizität im Vordergrund steht; hat die Funktion, die organische Form hervorzuheben	3D-Nachbildung des menschlichen Skeletts
Realnachbildung bzw. -modell	Ein Aspekt der Wirklichkeit eines untersuchten Vorgangs oder betrachteten Gegenstands wird originalgetreu, mitunter vergrößert, dreidimensional dargestellt, um den als wichtig erachteten Aspekt herauszustellen; Materialbeschaffenheit oft abweichend	3D-Kirschblütenmodell (100-fach vergrößert); im Schulfach Chemie mitunter auch 2D: Aluminiumwürfel als Modell für das auf der Kontinuumsebene (vgl. Kapitel 3.2.3 G, S 52) erfahrbare Al
Funktionsmodell	Technisches Modell, das „morphologisch stärker abstrahiert, hinsichtlich einer bestimmten Funktion“ (Schneider, 1981, S. 259), um diese Funktion herauszustellen.	Vergößerter Klappmechanismus bei Wiesensalbei Blüten
Simulation	Nachbildung von mitunter komplexen Prozessen als Denkmodell, Formeln und Computerprogramme; Simulationen können mitunter zur Emotionalisierung des Unterrichts genutzt werden (Simulationsspiele).	Nachspielen der neuronalen Reizweiterleitung in einer Schulklasse
Denkmodell	Denkmodelle sind Modelle bspw. von Prozessen, die nur gedanklich, also nicht weiter visualisiert,	In Gedanken nachvollzogene Wirkung von Neurotransmitter-

durchdacht werden.

hemmern

Anmerkung: Die Übersicht über die Modelltypen für den Biologieunterricht kann teilweise auch auf den Chemieunterricht (vgl. Tabelle 2) angewendet werden.

Für die folgende Modelldefinition aus der Chemie bzw. dem Chemieunterricht wurden auch Artikel aus der Physikdidaktik berücksichtigt, wenn es thematische Überschneidungen gab. Als ein Ergebnis aus dem systematischen Review lässt sich zeigen, dass in der chemiedidaktischen Literatur im Vergleich mit der biologiedidaktischen der Modellbegriff deutlich häufiger und einheitlicher definiert wird.

„Ein Modell ist ein Gegenstand oder theoretisches Konstrukt, das von einem Subjekt für einen entsprechenden Zweck geschaffen bzw. verwendet wird. Dabei bestehen zwischen bestimmten Entitäten des Modells und bestimmten Entitäten des präsentierten Objektes Analogien“ (Mikelskis-Seifert, Thiele & Wünscher, 2005, S. 35). Im Kontrast zu Mikelskis-Seifert et al. (2005) betonen andere Autoren das Lenken der Aufmerksamkeit auf spezifische Zielaspekte (Drechsler & van Driel, 2008, S. 612; Ingham & Gilbert, 1991, S. 193; Jong et al., 2005b, S. 950; Wang, Chi, Hu & Chen, 2014, S. 211). Damit werde es den Schülerinnen und Schülern möglich, abstrakte Zusammenhänge zu erklären und generalisierende Rückschlüsse daraus zu ziehen (Khourey-Bowers & Fenk, 2009, S. 442). Die Chemiedidaktiker van Driel und Verloop (1999, S. 1142) führen folgende Eigenschaften aus, die allen Modellen gemeinsam und somit für Schülerinnen und Schüler wichtig seien (Übersetzung: Beerenwinkel & Parchmann, 2008, S. 13):

1. „Ein Modell steht immer in Beziehung zu einem Original, welches durch das Modell repräsentiert wird.
2. Ein Modell ist ein Hilfsmittel des Wissenschaftlers, um Informationen über Originale zu erhalten, die nicht direkt beobachtbar oder messbar sind.
3. Ein Modell kann nicht direkt mit dem Original, das es repräsentiert, interagieren.
4. Ein Modell weist bestimmte Analogien zu dem Original auf.
5. Ein Modell unterscheidet sich immer in bestimmten Aspekten von dem Original.
6. Bei dem Entwurf eines Modells muss immer ein Kompromiss zwischen den Analogien und Unterschieden gefunden werden.
7. Die Entwicklung eines Modells ist ein iterativer Prozess.“

Im Folgenden werden einige wenige Modelltypen der Chemiedidaktik klassifiziert. Es fällt bei Auswertung des systematischen Reviews auf, dass hierüber weniger geschrieben wird als im Kontext biologiedidaktischer Forschung.¹²

Die Abgrenzung von Modell und Analogie ist bislang in der Chemiedidaktik weder scharf noch einheitlich. *Analogien* sind weit verbreitet bei der Vermittlung von Wissen. Neben sprachlichen gehören hierzu auch bildliche und spielerische Analogien (Harrison & Jong, 2005, S. 1). In Analogien im chemiedidaktischen Kontext kommt es zur „Übertragung von Strukturen oder Verfahrensweisen von einem Gegenstandsbereich auf einen anderen“ (Klinger, 2001, S. 170). Diese können Abbildungen enthalten, die nur teilweise gleich aussehen, teilweise jedoch davon abweichen. So können z. B. in der Modelldarstellung von Molekülen Form und Farbe nicht auf die realen kleinsten Teilchen übertragen werden (vgl. Mikelskis-Seifert et al., 2005, S. 35), weil auf der Diskontinuumsebene andere Zustände herrschen. Hierbei wird die enge und untrennbare Verwandtschaft der Begriffe Modell und Analogie deutlich, die auch die teilweise synonyme Verwendung rechtfertigen.

Mentale Modelle sind zum einen Repräsentationen von Zeichen, deren Bedeutung in der Auseinandersetzung mit der Umwelt entsteht und so im Gedächtnis gespeichert werden kann (vgl. Mikelskis-Seifert et al., 2005, 32 f.). In englischsprachigen Artikeln werden mentale Modelle zum anderen häufig als Repräsentationen von Präkonzepten verstanden. So behaupten Unal, Sadoglu und Durukan (2014, S. 675), dass innere Repräsentationen strukturelle Ähnlichkeiten zur erlebten Welt aufwiesen und diese personenspezifisch und unvereinbar mit wissenschaftlichen Vorstellungen seien. Der Unterschied zwischen den beiden Definitionen liegt in der Bewertung und im Geltungsbereich: Während die erste Definition jede Vorstellung, jedes „Wissen“ per se und im Sinne des Konstruktivismus als mentales Modell der Wirklichkeit versteht (vgl. Mikelskis-Seifert et al., 2005, 32 f.), wird unter der zweiten Definition nur ein nicht-wissenschaftliches Verständnis im Sinne eines Präkonzepts verstanden. Ein Überblick über Analogie und mentales Modell findet sich in Tabelle 2.

Tabelle 2: Ergänzender Überblick über Modelltypen im Chemieunterricht

Modelltyp	Definition	Beispiel
Analogie	Als Analogie wird die „Übertragung von Strukturen oder Verfahrensweisen von einem Gegenstandsbereich auf einen anderen“ (Klinger, 2001,	Sprachliche Übertragung des Begriffs Kette auf das diskontinuierliche Phänomen Molekülkette,

¹² Ingham und Gilbert (1991, S. 193) bilden eine Ausnahme mit ihrer Einteilung in maßstabsgetreue Modelle, Analogmodelle, mathematische, theoretische und archetypische Modelle. Da diese Kategorien teilweise Alleinstellungsmerkmal dieses Artikels sind, werden nur solche weiterverfolgt, die häufiger erwähnt werden.

	S. 170) bezeichnet.	z. B. bei Stärkemolekülen
mentales Modell (vgl. theoretische und Gedankenmodelle, Tabelle 1)	Zum einen Repräsentationen von Zeichen beim Denken, zum anderen Repräsentationen von Präkonzepten	Symbolische Übersetzung des Begriffs Wasser in das symbolische H ₂ O

Anmerkung: Die Übersicht über die Modelltypen für den Biologieunterricht (vgl. Tabelle 1) kann teilweise auch für den Chemieunterricht angewendet werden.

Der Vorgang des *Modellierens* (vgl. Kap. 3.2.3 I, S. 54) wird als der zentrale Vorgang naturwissenschaftlicher Schlussfolgerung verstanden und sollte für Schülerinnen und Schüler als zirkuläres Entwickeln und Überarbeiten von Modellen verständlich werden. Insofern ist die *Modellkompetenz* von Lernenden eng verknüpft mit dem Modellieren. Allerdings ist Modellkompetenz ein missverständlicher Begriff, da er sowohl auf Lernende als auch auf die Lehrkräfte bezogen werden kann. Wenn Upmeier zu Belzen und Krüger (2010, S. 49) die Modellkompetenz definieren als „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“, so beziehen sie sich auf Komponenten, die auf Lernende abzielen. Im Gegensatz zu Upmeier zu Belzen und Krüger (2010) beschreibt Meisert (2008, S. 245) die Dimensionen von Modellkompetenz als „Modellverständnis“ (als „übergeordnetes Konzept“, welches Schülerinnen und Schüler erlernen sollten), das sich aus dem „Modellwissen“ (Wissen um die (In-)Korrektheit von Modellen, Entstehung und Weiterentwicklung von Modellen und die Interpretationsbreite von Modellen) und der „Modellarbeit“ („Repertoire konkretisierter Beispiele“) speise. Wieder wird auf Schülerinnen und Schüler fokussiert; die Vermittlung von Modellkompetenz wird wiederum nur indirekt betrachtet.

Sind Schülerinnen und Schüler die Adressaten, so ist die Modellkompetenz ein Fachwissen. Betrifft Modellkompetenz Lehrkräfte, so wird sie zumindest von CK bzw. Wissen um die Natur der Naturwissenschaften (Nature of Science, kurz NoS, vgl. Kapitel 3.2.3 F: Wissen um den Unterschied verschiedener Modelltypen im Biologieunterricht, S. 52) beeinflusst. Darüber hinaus fehlt jedoch bislang eine genauere Erläuterung dessen, was *Elemente der Vermittlung von Modellkompetenz* sein könnten. Es bedarf einer selektiven Aufmerksamkeit bzw. des Noticings (vgl. Kapitel 3.2.3, S. 42) z. B. von Lernproblemen der Schülerinnen und Schüler bei der Entstehung von Modellkompetenz.

Zusammenfassend ergibt sich aus diesem Aspekt des Literaturreviews, dass

1. Modelle in Biologie- und Chemiedidaktik unterschiedlich definiert werden,
2. in biologiedidaktischen Artikeln seltener Modelle definiert werden und wenn doch, dann eher auf Modelltypen eingegangen wird und
3. in chemiedidaktischen Artikeln häufig und einheitlicher definiert wird, was Modelle sind, jedoch nur selten Modelltypen unterschieden werden.

3.2 Professionelle Unterrichtswahrnehmung

Lehrkompetenzmessung sollte sich am tatsächlichen Tun der Lehrkraft orientieren: am Unterrichten (vgl. Baumert & Kunter, 2006). Tests, die sich an der Unterrichtsrealität orientieren, sind rar und zum naturwissenschaftlichen Modelleinsatz nicht vorhanden. Es stellt sich die Frage, wie Unterricht simuliert werden kann und was je nach Art der Vignette damit gemessen wird. Unterrichtsvignetten werden zunehmend als Möglichkeit der Präsentation von Unterricht eingesetzt. Über das, was damit erfasst wird, wird diskutiert. Einige aktuelle Ansätze beschreiben Vignetten als Grundlage für *Professionelle Wahrnehmung*. Kapitel 3.2.1 (S. 38) beschreibt die Besonderheit *Professioneller Unterrichtswahrnehmung* generell als spezielle Kompetenz, während Kapitel 3.2.2 (S. 40) den Aspekt *Reasoning* genauer beschreibt und Kapitel 3.2.3 (S. 42) das *Noticing* von Modellkompetenz als weiteren Aspekt von *Professioneller Unterrichtswahrnehmung* ausdifferenziert.

3.2.1 Professionelle Unterrichtswahrnehmung als spezielle Kompetenz

Die Frage danach, was ein Test misst, ist die zentrale Frage der Validität. Aufgrund des besonderen Formats besteht beim Vignettentest seit Längerem eine Diskussion darüber, was dieser misst. Zwar kann seine Entstehung als Reaktion auf den Wunsch nach dem Messen von Kompetenzen verstanden werden (Kromrey & Renfrow, 1991), eine direkte Kompetenzerhebung ist jedoch unwahrscheinlich (Weinert, 2001). Fraglich sind die Kompetenzbestandteile, die Vignettentests erfassen können.

Nach Meschede (2014, S. 20) bestehen Zweifel daran, dass mit dem, was allgemein in der Fachdidaktik unter Professionswissen verstanden wird, die entscheidenden Aspekte des Lehrkräftehandelns beschrieben werden können; motivationale Orientierung, Überzeugungen bzw. Werthaltungen und Selbstregulation reichen nicht aus, um routinierte, flexible und adaptive Lehrkräftehandlung ausreichend zu beschreiben.

In Kapitel 2.3.2 (S. 27) wurden verschiedene Ansätze dessen, was Vignettentests testen, erläutert. Neben professioneller oder unterrichtlicher (Handlungs-)Kompetenz (Baer et al., 2007, S. 29; Brovelli et al., 2013, S. 307; Riese & Reinhold, 2010, S. 169), dem advokatorischen Ansatz (Oser et al., 2010, S. 8), neben dem im Klassenzimmer brauchbaren und nutzbaren Wissen (Kersting et al., 2012) und der Reflexionsfähigkeit (hinsichtlich Aufmerksamkeitsfokus, Bewertung und Analysefähigkeiten) (Schmelzing et al., 2010, S. 195) finden sich Operationalisierungsansätze wie „noticing and responding“ (Alonzo & Kim, 2012, S. 3), Wahrnehmung (perception) und Handlungsplanung (interpretation und decision-making) (PID-Modell: Blömeke et al., 2014) sowie „noticing and interpreting classroom interactions“ (van Es & Sherin, 2002, S. 572).

Zunehmend wird der ausdifferenzierte (vgl. Meschede, 2014, S. 13) Terminus „professional vision“ (Goodwin, 1994, S. 606) genutzt. Dieser stammt aus der Archäologie und definiert dort das aktive Wissen bei einer Ausgrabung, um gültige Schlussfolgerungen zu ziehen, was wiederum als Kodierungs-, Hervorhebungs- und Produktions- bzw. Artikulationsvorgang zusammengefasst wurde. Sherin (2007) sowie im Anschluss Sherin und van Es (2009, S. 22) wenden Professional Vision auf die Lehrtätigkeit an und fassen die Bestandteile als selektive Aufmerksamkeit („ability to notice“ (Sherin & van Es, 2009, S. 20)) und wissensbasiertes Schlussfolgern („knowledge-based reasoning“ (Sherin & van Es, 2009, S. 22)) zusammen. Im Deutschen wird in der Folge der Ausdruck *Professionelle Unterrichtswahrnehmung* (PU bzw. PUW, Seidel et al., 2010) geprägt. Auch in dieser Studie wird von diesem Konstrukt ausgegangen. Im Folgenden werden Reasoning und Noticing gesondert betrachtet.

Sherin (2007) beschreibt Professionelle Unterrichtswahrnehmung als einen von mehreren, aber einen entscheidenden Aspekt von Lehrkompetenz. Dass Professionelle Unterrichtswahrnehmung von Unterrichtsvignetten sowohl Einfluss auf das Unterrichten als auch auf den Lernerfolg von Schülerinnen und Schülern nimmt, können Kersting et al. (2012) beispielhaft zeigen.

Bei den bisherigen Definitionen dessen, was Professionelle Unterrichtswahrnehmung ausmacht, wurde zu wenig in Betracht gezogen, dass sich Professionalität bei der Unterrichtswahrnehmung auch durch die Adaptionsfähigkeit an bestimmte Themen auszeichnet. Für einen Vignettentest zur Vermittlung von Modellkompetenz ist das wissensbasierte Schlussfolgern bzw. Reasoning, aber vor allem das, was die selektive Aufmerksamkeit bzw. das Noticing ausmacht, entsprechend anzupassen. Die Suche nach einheitlichen Operationalisierungen

von Noticing und Reasoning kann somit nicht als sinnvoll erachtet werden, weil dies der Professionalität der Unterrichtswahrnehmung widerspricht.

3.2.2 Reasoning

Sherin und van Es (2009, S. 22) definieren Reasoning als die Art und Weise, wie eine Lehrkraft vor dem Hintergrund seiner auf Wissen und Verständnis gründenden Wahrnehmung (Noticing) beim Unterrichten wissensbasiert schlussfolgert. Diese zirkuläre Definition zeigt die enge Verknüpfung: Eine Schlussfolgerung kann nie ohne Wahrnehmung geschehen. Kleinknecht und Schneider (2013, S. 14) ergänzen, beim Reasoning ginge es um Reflexion und Interpretation von als entscheidend wahrgenommenen Unterrichtssituationen. Seidel und Stürmer (2014, S. 740) führen hohe Reasoning-Fähigkeiten auf ein differenziertes und integriertes Wissen zurück, das zu flexibler Anwendung in verschiedenen Lehr-Lern-Situationen führt. Sherin (2007, S. 385) weist auf die große Bedeutung des Fachwissens (CK) und des pädagogisch-psychologischen Wissens (PPK) hin, vor dessen Hintergrund Fragen wie „Warum hat die Schülerin diese Aussage so und nicht anders getroffen und was steht für ein Präkonzept hinter dieser Aussage?“ beantwortet werden müssten, um schlussfolgern und dann schließlich adäquat handeln zu können.

Über die Unterteilung von Reasoning besteht weitgehend Uneinigkeit. So wird Reasoning mit teilweise divergierenden Facetten beschrieben. Tabelle 3 zeigt einige dieser Unterteilungen.

Tabelle 3: Gegenüberstellung bisheriger Reasoning-Facetten

1) Seidel und Stürmer (2014, S. 745); 2) Kleinknecht und Schneider (2013, S. 14); 3) Kleinknecht (2014, S. 7)
 Stürmer und Seidel (2015, S. 55); (2013, S. 14);
 Kleinknecht und Schneider (2013, S. 14); Vgl. Sherin (2007, S. 390)

Beschreibung (description)	Zitieren (quotation)	Beschreiben
Erklärung (explanation)	Suche der Bedeutung (explore meaning)	Bewerten/ Erklären
Vorhersage (prediction)	Darstellung Schüleridee (synthesis of student ideas)	Schlussfolgern/ Alternativen entwickeln

Die Unterteilung von Spalte 1 und 3 ist weitgehend übereinstimmend, zumindest hinsichtlich der Facetten Beschreiben/Beschreibung und Erklärung/Erklären/Bewerten. Vorhersage von

Unterrichtsereignissen und Schlussfolgern/Alternativen entwickeln sind jedoch unterschiedliche Aspekte. Einen anderen Schwerpunkt legt Sherin (2007, S. 390), weil es ihr um einen sogenannten Videoclub¹³ und um die Unterscheidung von Noticing (hier noch als „selective attention“ bezeichnet) und Reasoning geht. Zitieren von Schülerinnen und Schülern, Suche nach der Bedeutung der Unterrichtssequenz und zusammenfassende Darstellung der Schüleridee können als komplexitätssteigernde Schlussfolgerungen verstanden werden, deren Vorstufe (keine Reaktion auf die Schüleridee) in der Aufteilung von Kleinknecht und Schneider (2013) bereits nicht mehr zitiert wird.

Meschede (2014, S. 18) verweist darauf, dass die Reasoning-Operationalisierungen je nach Arbeitsgruppe „teilweise sehr unterschiedlich“ seien. Hieraus ergibt sich die Freiheit, die Passung zum Erhebungsinstrument in den Vordergrund zu stellen. Für die vorliegende Studie wird nach der Literatursichtung eine andere Operationalisierung vorgenommen, weil die erwähnten Vorlagen zu divergierend sind und weil sie nicht zum hier entworfenen Vignettentest passen. Hier geht es bei Reasoning um die *Fähigkeit, aus*

1. *Diagnosemöglichkeiten,*
2. *Unterrichtsbewertungen und*
3. *Handlungsalternativen*

auswählen zu können. Entsprechend wurden alle Itemfragestellungen vor dem Hintergrund dieser Facetten gestellt. Die Auswahl zielt auf ein geschlossenes, sechstufiges Itemformat. Unter Diagnosemöglichkeiten wird das Diagnostizieren von Schülerpräkonzepten verstanden. Aufgrund der Annahmen bei der Testentstehung ist davon auszugehen, dass *unter Unterrichtsbewertungen auswählen können* schwieriger sein sollte als *aus Diagnosemöglichkeiten auszuwählen.* Am schwersten sollte das *Auswählen von Handlungsalternativen beim Beobachten von Unterricht* sein.

Entsprechend wurden für die Itembatterien Fragestellungen wie: „*Diagnostizieren* Sie die Schülervorstellung (Präkonzept) hinsichtlich des Immunsystems aus fachdidaktischer Sicht“, „*Bewerten* Sie den Unterrichtsverlauf aus fachdidaktischer Sicht“ oder zu Handlungsalternativen auswählen können beispielsweise: „Wie sollte der Lehrer den Unterricht fachdidaktisch geschickt weiterführen?“ konzipiert.

¹³ Videoclubs sind Seminarkonzepte, bei denen über das Betrachten und Reflektieren von Unterrichtsvideos versucht wird, handlungsnah Aspekte von Lehrkompetenz zu vermitteln Sherin (2007).

3.2.3 Noticing

Die Begriffe Noticing und selektive Aufmerksamkeit werden oft synonym verwendet. Selektive Aufmerksamkeit beschreibt die Entscheidung des bzw. der Beobachtenden, was er bzw. sie in einem bestimmten Moment beachten will (Sherin & van Es, 2009, S. 22). Entsprechend wird Noticing als Ermittlung entscheidender Unterrichtssituationen aus einer professionellen Perspektive verstanden (Stürmer & Seidel, 2015, S. 55). Stürmer und Seidel (2015, S. 55) beschreiben in ihrer Studie das Noticing als allgemeinpädagogische Wissensfacette (PK, Shulman, 1987) und setzen es mit den Prinzipien gelingenden Lehrens und Lernens gleich. Dementgegen wird Noticing in der vorliegenden Studie vor dem Hintergrund des Tests als fachdidaktisches Können bei der *Vermittlung von Modellkompetenz* in den Schulfächern Chemie und Biologie verstanden.

Zur Erhebung selektiver Aufmerksamkeit besteht eine Kontroverse. Zum einen wird davon ausgegangen, dass Noticing nur über Beobachtung wahrgenommen werden kann. In diesem Zusammenhang bieten sich Techniken wie das Eye-Tracking von Unterrichtssituationen an (vgl. z. B. Böttcher, Nitz & Upmeyer zu Belzen, 2015). Nachteil solcher Studien sind aufgrund des Aufwands die verhältnismäßig kleinen Stichprobengrößen und damit ihre beschränkte Aussagekraft auf die Grundgesamtheit.

Meschede, Steffensky, Wolters und Möller (2015, S. 320) gehen davon aus, dass selektive Aufmerksamkeit auch indirekt über offene und geratete Items erhoben werden könne, „da sie anders als geschlossene Formate den Fokus der Aufmerksamkeit nicht durch Itemtexte vorwegnehmen“. Über geschlossene Itemformate könne hingegen nur Reasoning erhoben werden, welches immer auch Elemente des Noticings enthalte (Meschede et al., 2015, S. 320).

Im Gegensatz dazu gehen wir für unseren Vignettentest davon aus, dass künftige Lehrkräfte aufgrund ihrer noch nicht ausreichenden Expertise durch geschlossene Itemformate zwar Hinweise auf Aufmerksamkeitsfokuse bekommen, diese jedoch erst mit steigender Expertise tatsächlich als Hinweis wahrnehmen können. Es besteht die Annahme für Lehramtsstudierende, dass es eine Expertiseschwelle gibt, die die Wahrnehmung von Hinweisen zu Aufmerksamkeitsfokussen voraussetzt. Hinweise für die Richtigkeit dieser Annahme werden im Validierungskapitel 6.2 (S. 79) begründet.

Modelle haben je nach Schulfach unterschiedliche Bedeutungen. Obwohl diese für die naturwissenschaftlichen Fächer sehr hoch ist, fehlt in den Naturwissenschaftsdidaktiken eine handlungsnahe Struktur und Auffächerung der Modellkompetenz von Lehrkräften in der Schule. Zwar legen Upmeyer zu Belzen und Krüger (2010, S. 53) einen Vorschlag zu „Struktur und

Niveaus der Modellkompetenz im Biologieunterricht“ vor. Dieser orientiert sich jedoch an den Schülerinnen und Schülern und nicht an den Lehrkräften. Ihre Einteilung bedient außerdem nicht die konkrete Modellthematik des Unterrichts, sondern hat eine Einteilung in Niveaustufen als Ziel. Mit den Bereichen „Kenntnisse über Modellbildung“ und „Modellbildung“ (Upmeier zu Belzen & Krüger, 2010, S. 53) sind zudem nicht alle Aspekte, die das Unterrichten von Modellen im Biologieunterricht (und Chemieunterricht) ausmachen, abgedeckt. Operationalisierungen sollten stets aus der Theorie hervorgehen. Das Thema „Modelle im Biologie- und Chemieunterricht“ wird in einer großen Menge an Artikeln aufgegriffen, ohne dass eine systematische Aufarbeitung der Forschungslage vorliegt. Um eine Operationalisierung des Begriffs Noticing zur *Vermittlung von Modellkompetenz* vornehmen zu können, erfolgt hier zunächst die Darstellung der Ergebnisse eines systematischen Literaturreviews in Form von Strukturdiagrammen und ergänzenden Beschreibungen. Artikel der Literaturdatenbanken „FIS Bildung“ und „Scopus“ wurden berücksichtigt und nach den Begriffen Naturwissenschaften, Didaktik, Modell, Unterricht, Fachdidaktik, Chemie, Biologie bzw. science, PCK (Pedagogical Content Knowledge), model(s), teaching, chemistry, biology durchsucht. Die Artikel, welche sowohl die Lehrkräfte als auch die Schülerinnen und Schüler berücksichtigen, wurden ausgewählt, um aus beiden Perspektiven Rückschlüsse auf adäquates Unterrichten von Modellen zu ziehen. Über die Suchbegriffe wurden 543 Artikel (160 in FIS, 383 in Scopus) gefunden.

Depaepe, Verschaffel und Kelchtermans (2013, 14 f.) selektieren Artikel nach fünf Kriterien: Erstens müssen die Artikel aus peer-reviewten Zeitschriften stammen, zweitens müssen sie in englischer Sprache verfasst sein, drittens müssen sich die Artikel tatsächlich mit dem Thema (hier pedagogical content knowledge (PCK)) und nicht mit verwandten Begriffen (z. B. technological pedagogical and content knowledge (TPACK)) auseinandersetzen, viertens müssen sich die Artikel fachspezifisch mit dem Thema befassen (hier: Mathematik und nicht beispielsweise Naturwissenschaften) und fünftens müssen die Artikel empirisch (qualitativ bzw. quantitativ) und nicht anderer Art sein (z. B. konzeptuelle oder Diskussionsbeiträge).

Angelehnt an dieses Vorgehen wurden Artikel entfernt. Allerdings müssen die Artikel in diesem systematischen Literaturreview nicht aus peer-reviewten Zeitschriften stammen, da Erkenntnisse über die *Elemente zur Vermittlung von Modellkompetenz* auch aus anderen Quellen zu erwarten sind und die Aussagekraft hier nicht abhängig sein muss von den Selektionskriterien, über die die Artikel verlegt wurden. Neben englischen wurden auch deutschsprachige Artikel in das Review aufgenommen. Artikel, die sich mit der Entwicklung von empirischen Modellen auseinandersetzen und dies nicht hinsichtlich von Modellen im Unterricht

tun, wurden ebenso entfernt wie Artikel, die sich nicht auf die naturwissenschaftlichen Schulfächer Chemie und Biologie bezogen. Abweichend von Depaepe et al. (2013, S. 15) wurden konzeptuelle, theoretische oder Diskussionsbeiträge nicht entfernt, weil aus solchen ebenso Hinweise auf *Elemente zur Vermittlung von Modellkompetenz* gezogen werden können.

Nach der Entfernung verblieben 57 Artikel (29 bei Scopus, 28 bei FIS). Enthalten sind auch Artikel, die sich mit der Abgrenzung von Modellen und Analogien, mentalen Modellen und Repräsentationen und damit implizit mit Modellen beschäftigen. In diesem Fall wurde nur der Bereich Modell weiter ausgewertet.

Depaepe et al. (2013, S. 15) wertet Artikel nach folgenden neun Kriterien aus:

1. die Definition (hier von PCK),
2. die Komponenten, die das Konstrukt ausmachen (hier von PCK),
3. die Forschungsfrage,
4. die Forschungsmethode,
5. die fachwissenschaftliche Domäne, aus der das Konstrukt untersucht wurde (hier Bereich der Mathematik),
6. das Land, in dem die Studie durchgeführt wurde,
7. die Stichprobengröße,
8. die Art der Probandinnen- und Probanden und
9. die Forschungsergebnisse.

Aus dieser Aufstellung wurden für diese Arbeit folgende neun Analysekriterien abgeleitet und auf die verbliebenen Artikel angewendet:

1. Art des Artikels (normativ, theoretisch, empirisch: qualitativ, quantitativ),
2. Explizite Definition des Modellbegriffs,
3. Elemente der Vermittlung von Modellkompetenz,
4. Land der Studie bzw. Ort der Datenerhebung,
5. Forschungsfrage,
6. Auswertungsmethode,
7. Stichprobengröße,
8. Population der Befragten,
9. zentrale Forschungsergebnisse.

Fachwissenschaftliche bzw. -didaktische Domäne ist stets die Chemie und Biologie. An dieser Stelle wird nur von der Analyse der Artikel hinsichtlich der Zusammenstellung der *Elemente zur Vermittlung von Modellkompetenz* berichtet, die hier zugleich als Noticing bezeichnet

net werden. Die Modellkompetenz, die eine Biologielehrkraft besitzen sollte, um lernförderlich mit Modellen zu unterrichten, ist in Abbildung 4 operationalisiert, jene, die eine Chemielehrkraft haben sollte, in Abbildung 5. Die Strukturdiagramme enthalten die zentralen Aspekte der Modellvermittlung und können Grundlage empirischer Erprobung sein. Die Anzahl der Nennungen in den Artikeln wird in Klammern hinter den *Elementen zur Vermittlung von Modellkompetenz* erwähnt, um einerseits Hinweise über deren Wichtigkeit anzudeuten. Andererseits gab es eine Tendenz, häufig genannte *Elemente zur Vermittlung von Modellkompetenz* bei der Entwicklung des Vignettentests eher zu berücksichtigen.

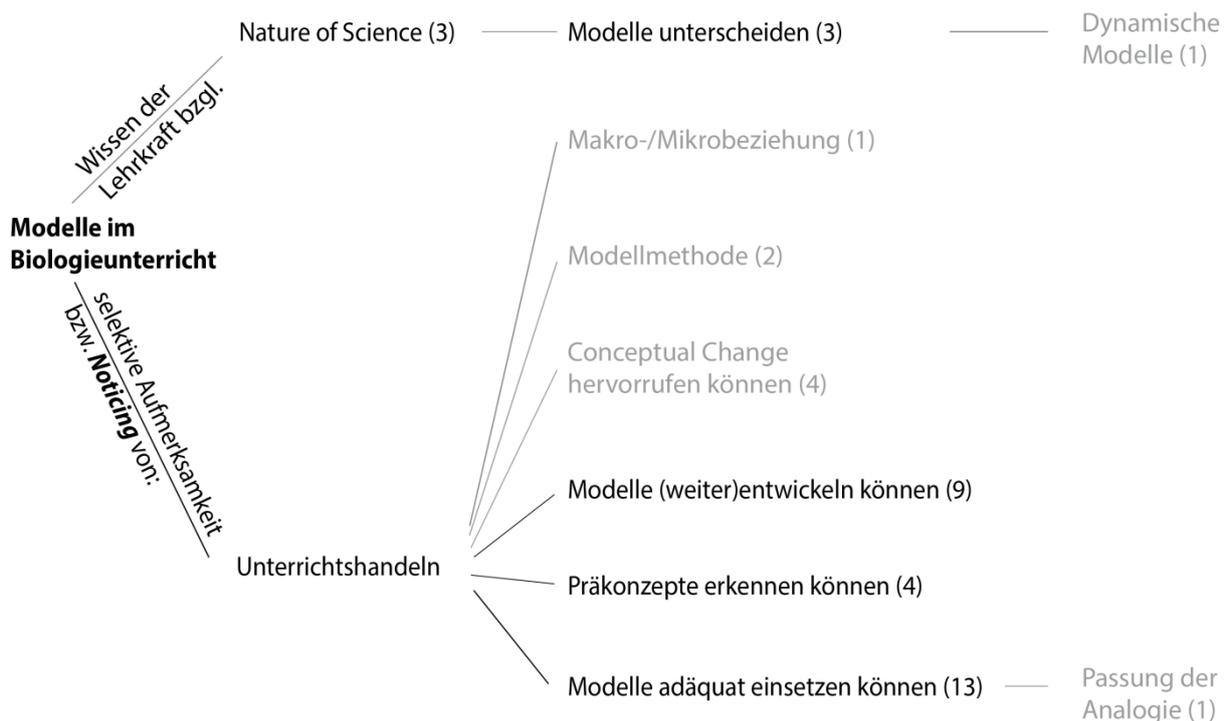


Abbildung 4: Strukturdiagramm zur Operationalisierung der *Elemente zur Vermittlung von Modellkompetenz* (Noticing) im Biologieunterricht. In Klammern findet sich die Anzahl der Artikel, die den jeweiligen Aspekt benennen. Hellgrau sind all jene Facetten, die bei der Entwicklung des Vignettentests unberücksichtigt bleiben mussten, eigene Darstellung.

In Anlehnung an Upmeier zu Belzen und Krüger (2010, S. 53) hat auch dieses Modell zur Modellkompetenzvermittlung für den Biologieunterricht in Abbildung 4 eine Wissenskomponente und eine Handlungskomponente. Der *adäquate Modelleinsatz*, beispielsweise hinsichtlich der Passung des Modells zur Aufgabe im Unterricht, wurde am häufigsten genannt. Am zweithäufigsten wurde in den Artikeln das *Entwickeln und Weiterentwickeln von Modellen* erwähnt. Damit ist sowohl die Fähigkeit der Lehrkraft gemeint, in der Vorbereitung des Unterrichts Modelle zu entwickeln bzw. vorhandene weiterzuentwickeln, als auch Schülerinnen und Schüler im Unterricht darin zu unterstützen. Viermal wurde in den Artikeln die Notwen-

digkeit genannt, beim Einsatz von Modellen im Unterricht an die *Präkonzepte* der Schülerinnen und Schüler anzuknüpfen.

Nicht berücksichtigt wird in dieser Studie das Wissen um *dynamische Modelle* (Crawford & Cullin, 2004) und wie diese sich von anderen Modellen unterscheiden. Zum einen sind dynamische Modelle äußerst heterogen und es ist somit schwierig, ein passendes exemplarisch auszuwählen. Zum anderen ist das hinter dynamischen Modellen stehende Konstrukt noch zu wenig präzise, um es für diesen Test operationalisieren zu können. Obwohl viermal genannt, wird *Conceptual Change hervorrufen können* (Cohen & Yarden, 2009; Crawford & Cullin, 2004; vgl. Posner, Strike, Hewson & Gertzog, 1982; Valanides & Angeli, 2006; van Dijk, 2009) bei der Entwicklung des Biologietests nicht explizit berücksichtigt. Da das Problem der Schülerinnen- und Schülerpräkonzepte immer auch nach deren Lösung im Conceptual Change verlangt, wird es jedoch implizit mit einbezogen. Das Konzept des Conceptual Changes hat sich außerdem nach Posner et al. (1982) eher in der Chemie- und Physikdidaktik durchgesetzt, während in der Biologiedidaktik vorsichtiger von Conceptual Reconstruction (vgl. van Dijk & Kattmann, 2007) die Rede ist. Die *Modellmethode* (Leisner-Bodenthin, 2006; vgl. Nelson & Davis, 2012; vgl. Valanides & Angeli, 2006) findet sich ebenfalls nicht in der Umsetzung des Vignettentests Biologie, wie auch nicht die *Makro-Mirko-Beziehung* (Cohen & Yarden, 2009). Grund hierfür ist zum einen, dass es sich um außergewöhnliche Nennungen handelt: Insbesondere die Berücksichtigung der Makro-Mikro-Beziehung in der Biologiedidaktik kann als Alleinstellungsmerkmal bezeichnet werden. Die Modellmethode findet sich eher in Artikeln aus dem Kontext der Chemie- und Physikdidaktik und wird deshalb dort berücksichtigt.

Auch im Ergebnis der aus dem systematischen Literaturreview gewonnenen *Elemente zur Vermittlung von Modellkompetenz* im Chemieunterricht (vgl. Abbildung 5) finden sich die beiden Komponenten *Handeln* und *Wissen*. Allerdings ist das Wissen hier differenzierend ergänzt um das *Curriculum* (Justi & van Driel, 2005b, 2005c) der Schule hinsichtlich Modellen sowie um die NoS-Aspekte *Wissen über Modelle in der Geschichte der Chemie* (Beerenwinkel & Parchmann, 2008; Bindernagel & Eilks, 2009; Christen, 1994; Crawford & Cullin, 2004; Garritz, 2013; Graulich, Schreiner & Hopf, 2010; Gröger, 2010; Hammer, 1988; Justi & van Driel, 2005c), *philosophisches Wissen über Modelle* (Crawford & Cullin, 2004; Garritz, 2013; Padilla, Ponce-de-León, Rembado & Garritz, 2008; Šorgo et al., 2014) und *Modelle unterscheiden* (Crawford & Cullin, 2004; Schmelzing et al., 2013b). Diese Wissensaspekte wurden jedoch bei der Entwicklung des Chemietests nicht berücksichtigt, insbesondere um die Testzeit in Chemie zu reduzieren. Da es weniger Chemie- als Biologielehramtsstudierende

gibt, musste hier besonders darauf geachtet werden, dass möglichst wenige Probandinnen und Probanden den Test abbrechen. Insofern stellt das *Wissen über Modelle* ein in diesem Test unbearbeitetes Desiderat dar.

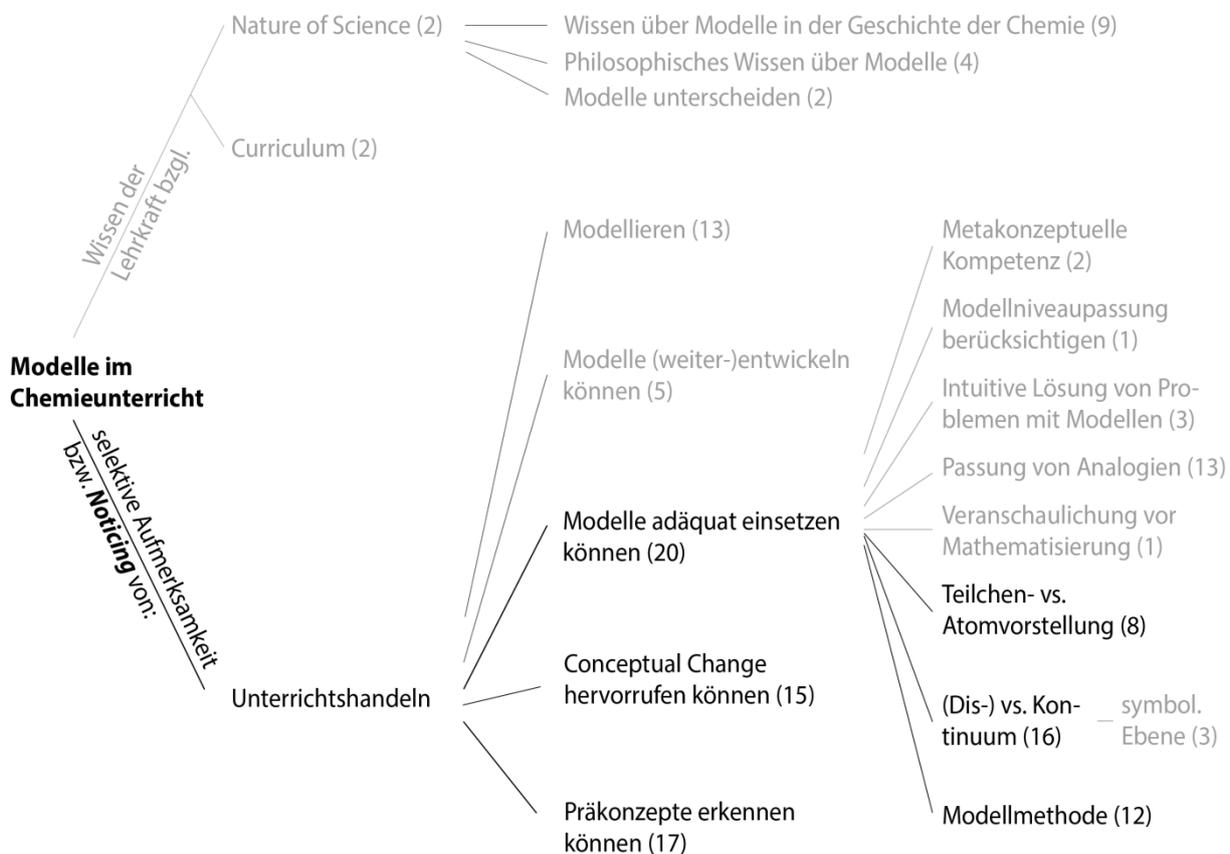


Abbildung 5: Strukturdiagramm zur Operationalisierung der *Elemente zur Vermittlung von Modellkompetenz* (Noticing) im Chemieunterricht. In Klammern findet sich die Anzahl der Artikel, die den jeweiligen Aspekt benennen. Hellgrau sind all jene Facetten, die bei der Entwicklung des Vignettestests nicht berücksichtigt werden konnten, eigene Darstellung.

Bei der Operationalisierung (vgl. Abbildung 5) werden berücksichtigt: die Unterrichtshandlungsaspekte *Präkonzepte erkennen können*, *Conceptual Change hervorrufen können* und *Modelle adäquat einsetzen können*. Letzteres teilt sich in folgende Unterfacetten auf: *Teilchen- vs. Atomvorstellungen*, *Diskontinuum vs. Kontinuum* (ohne dabei die von Johnstone (1991) vorgeschlagene symbolische Dimension explizit zu berücksichtigen, implizit ist sie ebenfalls enthalten) und *Modellmethode*. Zum *adäquaten Modelleinsatz* gehören außerdem die nicht berücksichtigten Unterfacetten *metakonzeptuelle Kompetenz* (Leisner & Mikelskis, 2004; Mikelskis-Seifert, 2004), die Berücksichtigung der *Modellniveaupassung* (Christen, 1994), die Förderung von Schülerinnen und Schülern bei der *intuitiven Lösung bei Problemen mit Modellen* (vgl. Garritz, 2013, S. 1801; Graulich et al., 2010; vgl. Padilla et al., 2008, S. 1396), die *Passung von Analogien* zum Unterricht (Beerenwinkel & Parchmann, 2008;

Eilks, Witteck & Pietzner, 2009a; Gröger, 2010; Harrison & Jong, 2005; Khourey-Bowers & Fenk, 2009; Klinger, 2001, 2001; Mikelskis-Seifert, 2004; Mikelskis-Seifert et al., 2005; Oh & Kim, 2013; Padilla et al., 2008; Valanides & Angeli, 2006; Wang et al., 2014) und das Prinzip *Veranschaulichung vor Mathematisierung* (Klinger, 2001), welches der Physikdidaktik entstammt, sich aber auf die Chemiedidaktik übertragen lässt. Außerdem wurde die Unterrichtshandlungsfacette *Modellieren* (Jong et al., 2005b; Justi & van Driel, 2005a, 2005b, 2005c; Leisner & Mikelskis, 2004; Lethmate & Arning, 2003; Mikelskis-Seifert et al., 2005; Mikelskis-Seifert & Leisner, 2004; Nelson & Davis, 2012; Strübe, Tröger, Tepner & Sumfleth, 2014; Unal et al., 2014; Valanides & Angeli, 2006; Wang et al., 2014) nicht explizit erfasst, da diese als Handlungsaspekt des Modelleinsatzes nicht hinreichend abgegrenzt ist und somit keine zusätzlichen Informationen bietet, im Gegensatz zur Mathematikdidaktik, bei der der Modellierungsprozess exakt definiert ist (vgl. z. B. Blum et al., 2002). Schließlich wurde der Aspekt *Modelle (weiter-)entwickeln können* (König & Reiners, 2004; Leisner & Mikelskis, 2004; Mikelskis-Seifert, 2004; Mikelskis-Seifert et al., 2005; Streller & Bolte, 2008) nicht explizit weiterverfolgt, weil in der Chemiedidaktik andere Aspekte stärker konkretisiert wurden und dadurch exakter operationalisierbar waren.

Im Folgenden werden die im Biologie- und/oder Chemietest berücksichtigten Operationalisierungen vorgestellt, definiert und in den Bedeutungskontext des als Grundlage genutzten systematischen Literaturreviews gestellt. Es werden die Aspekte *Modelle adäquat einsetzen können*, *Modelle (weiter-)entwickeln können*, *Präkonzepte erkennen können*, *Conceptual Change hervorrufen können*, Wissen, dass und wie sich *Modelle unterscheiden*, *Modellmethode*, *Realmodelle vs. Modelle des Diskontinuums* (im Folgenden (*Dis-*) vs. *Kontinuum* genannt) und *Teilchen- vs. Atomvorstellungen* vorgestellt. Damit wird auch der Rahmen des Noticings hinsichtlich *Elemente zur Vermittlung von Modellkompetenz* beschrieben.

A) Modelle adäquat einsetzen können

Modelle adäquat einsetzen können wird im Folgenden als Überkategorie verwendet. 13 Artikel zu den biologischen Modellen beziehen sich auf diese Kategorie (Cohen & Yarden, 2009; Crawford & Cullin, 2004; Gad & Mittelsten Scheid, 2008; Hilgers, Sommerfeld & Blume, 1994; Lethmate & Arning, 2003; Markert, 1988; Meyer, 1990; Nocke, 2001; Romine & Walter, 2014; Schneider, 1981; Šorgo et al., 2014; Steinbeck & Markert, 1989; Valanides & Angeli, 2006). Bei den chemischen Modellen im Unterricht beziehen sich 20 Artikel auf den adäquaten Modelleinsatz (Cohen & Yarden, 2009; Drechsler & van Driel, 2008; Harrison &

Jong, 2005; Justi & van Driel, 2005b, 2005c; Klinger, 2001; König & Reiners, 2004; Lethmate & Arning, 2003; Mikelskis-Seifert, 2004; Mikelskis-Seifert et al., 2005; Mikelskis-Seifert & Leisner, 2004; Okanlawon, 2010; Padilla et al., 2008; Papageorgiou, Stamovlasis & Johnson, 2010; Rueckl & Ebinghaus, 1989; Streller & Bolte, 2008; Strübe et al., 2014; Touché, 1989; Usak, Ozden & Eilks, 2011; Winnenburg, 2000). Zwar betonen Henze, van Driel und Verloop (2007, S. 105) die Relevanz des adäquaten Verstehens der Natur der Modelle und des Modellierens für den Chemieunterricht. Die Kategorie *Modelle adäquat einsetzen können* geht jedoch darüber hinaus, da es sich dabei im Kern nicht um eine Fachwissens- oder Nature-of-Science-Kategorie, sondern um eine fachdidaktische Überkategorie handelt. Zugleich wird *Modelle adäquat einsetzen können* auch als Ausweichkategorie verwendet, wenn eine spezifische Auseinandersetzung mit den *Elementen zur Vermittlung von Modellkompetenz* nicht zielführend für die Ausrichtung des Artikels war. Hieraus folgt, dass eine übergreifende Definition von *Modelle adäquat einsetzen können* nicht erfolgen kann. Vielmehr wird es differenziert in die folgenden Unterkategorien (vgl. S. 51 ff.): *Atommodelle vs. Teilchenvorstellungen, (Dis-) vs. Kontinuum* und *Modellmethode*.

B) Modelle (weiter)entwickeln können

Dass Modelle im Wissenschaftsbetrieb ständig entwickelt und weiterentwickelt werden, sollte Teil des Wissens um die Natur der Naturwissenschaften (NoS) oder des Fachwissens von Lehrkräften sein. Obwohl diese Kategorie präziser ist als die Kategorie *Modelle adäquat einsetzen können*, bleibt sie relativ unspezifisch, da sie das Wissen über diesen Vorgang in der Wissenschaft, die Fähigkeit der Lehrkraft, Modelle für den Unterricht (weiter) zu entwickeln und die Kompetenz, Schülerinnen und Schülern eine adäquate Entwicklung und Weiterentwicklung von Modellen zu vermitteln, umfasst. Mikelskis-Seifert (2004, S. 20) versteht unter dieser Kategorie die Kompetenz, ein „Metakonzept zur Teilchenvorstellung zu entwickeln“. Keune und Daemmgen (1988) regen den Nachbau von Glucose-Modellen mit Zeichenkarton an. Schneider (1981) schlägt den Einsatz quantifizierender Modelle zur Wirbelsäule vor und Nelson und Davis (2012, S. 1932) eine Bewertung von Modellqualität von Schülerinnen und Schülern durch Berücksichtigung von Klarheit, Bedeutung und Übereinstimmung mit empirischen Belegen. Die anderen neun Artikel (Crawford & Cullin, 2004; Gad & Mittelsten Scheid, 2008; Hilgers et al., 1994; König & Reiners, 2004; Leisner & Mikelskis, 2004; Litsche, 1985; Meyer, 1990; Mikelskis-Seifert et al., 2005; Mikelskis-Seifert & Leisner, 2004) gehen unspezifischer mit der Thematik um.

C) Präkonzepte erkennen können

Präkonzepte werden in der Literatur teilweise als Miss-, Fehl- oder naive Konzepte bezeichnet (vgl. hierzu als Überblick z. B. Lin & Chiu, 2010, S. 1618). Andere Autorinnen und Autoren formulieren bewusst wertfrei und definieren Präkonzepte als vorwissenschaftliche Alltagsvorstellungen von Lernenden (vgl. Parchmann, 2012), die von den wissenschaftlichen Erkenntnissen abweichen. Der Umgang mit Präkonzepten bildet jedoch den entscheidenden Schlüssel zum nachhaltigen Lernen von Schülerinnen und Schülern. Entsprechend betonen viele Studien die positive Wirkung eines adäquaten Umgangs der Lehrkraft mit Präkonzepten auf das Lernen der Schülerinnen und Schüler. So auch beim Ziel, Modellkompetenz zu vermitteln (Coll & Treagust, 2003; Park & Light, 2009).

Präkonzepte von Schülerinnen und Schülern betreffen häufig Modelle, die im Chemie- und Biologieunterricht eingesetzt werden. So konnten Lin und Chiu (2010) zeigen, dass sich die Präkonzepte zu Modellen hinsichtlich Säuren und Basen leistungsschwacher Schülerinnen und Schüler nach einer Unterrichtseinheit verstärkten, während leistungsstarke Schülerinnen und Schüler ihre Präkonzepte in wissenschaftliche Vorstellungen überführten. Obwohl es sich um eine qualitative Studie handelt, stellt sich die Frage, ob Leistungsstärke auch daraus resultiert, dass Schülerinnen und Schüler besser mit den eingesetzten Analogien, Grafiken etc. zurechtkommen. Im Umkehrschluss gilt es die Frage zu beantworten, wie insbesondere auf die Präkonzepte leistungsschwacher Schülerinnen und Schüler adäquat reagiert werden könnte. Lin und Chiu (2010, S. 1618) schlagen eine Sammlung von Präkonzepten für Lehrkräfte vor, die auch Hinweise auf ein angemessenes Vorgehen hinsichtlich Conceptual Change enthalten.

Für taiwanesischen Lehrkräfte wurde eine solche Sammlung entwickelt (vgl. Vorarbeiten im *International Journal of Science Education*, 2007, Vol. 29 (4)). International ist jedoch bislang weder eine umfassende Sammlung von Präkonzepten für die Naturwissenschaftsdidaktiken noch eine auf Wirksamkeit überprüfte Liste mit adäquaten Reaktionen im Unterricht erschienen. Eine konsequente Unterrichtsplanung mit wissenschaftlich ermittelten Präkonzepten ist zumindest in Deutschland noch nicht üblich (vgl. Schrader, 2015). Für den Biologieunterricht sei auf die Präkonzeptsammlung von Kattmann (2015) verwiesen, sowie beispielsweise auf Trauschke (2008, S. 495) und Cohen und Yarden (2009, S. 145).

Für den Chemieunterricht geben folgende Reviews einen Überblick über Präkonzepte: Garritz (2013, S. 1790), Eilks et al. (2009b, 53 und 58), Jong und van Driel (2004, S. 480), Liang, Chou und Chiu (2011, S. 238), Okanlawon (2010, S. 34).

D) *Conceptual Change*

Die Frage nach der Überwindung von Präkonzepten beschäftigt die Didaktik wahrscheinlich schon seit Beginn des Lehrens. Diese kann als Assimilation bzw. evolutionärer Vorgang durch vorsichtige Angleichung oder als Akkomodation bzw. revolutionärer Vorgang durch radikale Änderung der vorwissenschaftlichen an wissenschaftliche Vorstellungen erfolgen (Posner et al., 1982, S. 212). Nach Posner et al. (1982, S. 214) geschieht die Akkomodation von Präkonzepten (auch bezüglich Modellen) über folgende vier Voraussetzungen:

1. Dissatisfaction: Eine gewisse Unzufriedenheit muss bei Lernenden vorhanden sein.
2. Intelligible: Die Lernenden müssen die wissenschaftliche Vorstellung logisch finden.
3. Plausible: Die Lernenden müssen die wissenschaftliche Vorstellung als überzeugend empfinden.
4. Fruitful: Die Lernenden müssen die wissenschaftliche Vorstellung als fruchtbar empfinden (vgl. Duit, 2000, S. 81; Posner et al., 1982, S. 214).

Am Konzept des *Conceptual Change* hat sich seit den 1980er-Jahren wenig verändert. Heute wird der Begriff *Conceptual Change* vorsichtiger benutzt (Duit & Treagust, 2012, 49 f.) und eher als Rekonstruktion (vgl. van Dijk & Kattmann, 2007) verstanden.

Zwar wird auch im Kontext der Biologie *Conceptual Change* bzw. *Konzept-Rekonstruktion* thematisiert. Allerdings finden sich hier nur vier Studien (Cohen & Yarden, 2009; Crawford & Cullin, 2004; Valanides & Angeli, 2006; van Dijk, 2009) im systematischen Literaturreview. Deutlich häufiger, nämlich 15 mal, finden sich Artikel aus dem Kontext der Chemie (Beerenwinkel & Parchmann, 2008; Benedict & Bolte, 2009; Eilks et al., 2009a; Garritz, 2013; Harrison & Jong, 2005; Khourey-Bowers & Fenk, 2009, S. 441; Leisner & Mikelskis, 2004; Liang et al., 2011, S. 238, 2011; Lin & Chiu, 2010, S. 1618; Mikelskis-Seifert, 2004; Mikelskis-Seifert et al., 2005; Mikelskis-Seifert & Leisner, 2004; Papageorgiou et al., 2010; Winnenburg, 2000).

E) *Modellmethode*

Als *Modellmethode* wird hier der von Leisner-Bodenthin (2006, S. 98) erläuterte Unterrichtsablauf mit folgenden Elementen verstanden:

1. Phänomenbeobachtung: das Phänomen ist mithilfe des „Originals“ nicht erklärbar;
2. „Modellentwicklung“ bzw. „Modellauswahl“;
3. „Beantworten der Frage“ bzw. „Lösen des Problems (Modellanwendung)“;

4. „Zweckmäßigkeit und Erklärungswert des Modells prüfen – Reflexion der Modellnutzung“.

Der Vorteil einer Orientierung des Unterrichts an einem solchen Schema liegt insbesondere im vierten Schritt, nämlich dem Hinterfragen des Modells durch die Schülerinnen und Schüler. Wenn die Schülerinnen und Schüler die *Modellmethode* verinnerlicht haben, können sie Modelle als zweckorientierte Konstrukte erkennen, die nicht immer wahr, sondern durchaus fehlerbehaftet sein können. Im vorgenommenen Review erscheint die *Modellmethode* in zwölf chemiedidaktischen Artikeln (Eilks et al., 2009a, S. 54; Harrison & Jong, 2005, S. 2; Justi & van Driel, 2005b, 2005c; Leisner & Mikelskis, 2004; Lethmate & Arning, 2003; Mikelskis-Seifert, 2004; Mikelskis-Seifert et al., 2005; Mikelskis-Seifert & Leisner, 2004; Nelson & Davis, 2012, S. 1932; Strübe et al., 2014; Valanides & Angeli, 2006).

F) Wissen um den Unterschied verschiedener Modelltypen

Da Modellwissen einen wichtigen Einfluss auf die Modellkompetenz hat (vgl. z. B. Justi & van Driel, 2005c), ist anzunehmen, dass es diesen auch auf die Fähigkeit hat, Modellkompetenz zu vermitteln. Modellwissen wird hier im Sinne von Meisert (2008, S. 244) als „Kenntnis grundlegender Modelle“ und als Fähigkeit, diese unterscheiden zu können, verstanden. Modellwissen resultiert bei Lehrkräften zum einen aus ihrem Fachwissen, zum anderen aus dem Wissen um die Natur der Naturwissenschaften (Nature of Science, NoS). Hierzu zählt auch das Unterscheiden, z. B. von *Modellen und Nicht-Modellen* (Meisert, 2008, S. 246) und von *Struktur- und Funktionsmodellen* (Meisert, 2008, S. 256). Auch Garritz (2013, S. 1788) betont die Bedeutung der Unterscheidungsfähigkeit von *Analogien* und Modellen für ein adäquates Verständnis von NoS. Hieran schließen Šorgo et al. (2014, S. 6) an und berichten länderübergreifend (Tschechien, Slowakei, Slowenien, Türkei) von großen Wissenslücken bei Erstsemestern des naturwissenschaftlichen Lehramts bezüglich NoS im Bereich Genetik, Evolution und Menschwerdung, auch hinsichtlich von Modellen zu diesen Themen. Zur Steigerung des Wissens über die Natur der Naturwissenschaften hinsichtlich des Modellwissens schlagen Crawford und Cullin (2004, S. 1386) Seminare explizit zum Wissen über Modelle vor und zeigen, dass auch Modelliersoftware wie Model-IT hierzu dienlich sein kann.

G) Mit diskontinuierlichen Modellen und Realmodellen umgehen können

Eine häufig auftretende Schwierigkeit bei der Vermittlung z. B. von Atommodellen im Chemie- aber auch im Physikunterricht, ist die Unterscheidung der thematisierten Ebenen. Im

Unterricht wird das *Phänomen*, d. h. die makro-, mikroskopische oder die Kontinuumsebene, häufig vermennt mit der *Atomebene* (mikro-, submikroskopische oder Diskontinuumsebene). Die Unterscheidung dieser beiden Ebenen ist jedoch entscheidend, da auf der Diskontinuums-ebene andere Naturgesetze herrschen. Für Lernende verschwimmen während des Unterrichts oft die Grenzen (Rehm, 2007). So vermischen sich Kontinuums- mit Diskontinuumsvorstellungen zu einer unwissenschaftlichen Vorstellung, zu einem von der Schule geprägten Präkonzept. Cohen und Yarden (2009) stellen ähnliche Probleme beim Verständnis des mikroskopischen Bildes im Biologieunterricht fest. Dennoch bleibt die Vermischung der Kontinuums- und Diskontinuumsvorstellungen hauptsächlich ein chemie- und physikdidaktisches Problem. Johnstone (1991) erkennt in der Repräsentation solcher Modelle ein weiteres Problem und ergänzt die symbolische Ebene, die sowohl Makro- als auch Mikroebene enthält. Jong und van Driel (2004) identifizieren beim Lehren von Modellen im Chemieunterricht folgende Lehrschwierigkeiten:

1. Zu schnelles Schlussfolgern von der Mikro- zur Makrobedeutung (S. 483),
2. Hauptsächliche Orientierung auf die Mikrobedeutung (S. 485),
3. Vermischung von Mikrobedeutung und Makrobedeutung (S. 485),
4. Umgang mit Vorschriften zur Notierung von symbolischen Repräsentationen (S. 486).

Jong und van Driel (2004, S. 477) können mit ihren Forschungsergebnissen zu künftigen Chemielehrkräften zeigen, dass sich durch Seminare zum adäquaten Unterrichten von Modellen das Wissen um Lehrschwierigkeiten hinsichtlich der oben erwähnten vier Bereiche steigern lässt. Die Bedeutung des Umgangs mit dem (Dis-)Kontinuums-Problem spiegelt sich auch in der hohen Anzahl von 16 Artikeln im Review wider (Benedict & Bolte, 2009; Bühler & Erb, 2010; Drechsler & van Driel, 2008; Eilks et al., 2009a; Garritz, 2013; Harrison & Jong, 2005; J. H. van Driel, O. d. Jong & N. Verloop, 2002; Jong & van Driel, 2004; Justi & van Driel, 2005a; Khourey-Bowers & Fenk, 2009; Liang et al., 2011; Lin & Chiu, 2010; Mikelskis-Seifert, 2004; Mikelskis-Seifert et al., 2005; Mikelskis-Seifert & Leisner, 2004; Wang et al., 2014).

H) Mit Teilchen und Atomen umgehen können

Atome und Teilchen sind zentraler Bestandteil des Chemie- und Physikunterrichts. Während Atome weitgehend einheitlich definiert werden, werden unter Teilchen bzw. Partikeln sowohl Bestandteile von Atomen als auch Bestandteile von Materie allgemein verstanden. Schon dies kann Schülerinnen und Schülern Schwierigkeiten verursachen. Die Abhängigkeit einer ange-

messenen Benutzung des Teilchenmodells bzw. des Atommodells von der Thematik und der Altersstufe ist schon häufig diskutiert worden (Rehm, 2007). Allerdings geben auch die im Rahmen des Literaturreviews untersuchten zwei Artikel (Mikelskis-Seifert et al., 2005; Papa-georgiou et al., 2010) hierzu keine eindeutige Auskunft. Touchés (1989) Klärung der Begrifflichkeiten für die Unterrichtsfächer Chemie und Physik leistet hierzu einen wichtigen Beitrag.

1) Modellieren vermitteln können

Aufgrund der Bedeutung der Vermittlung des *Modellierens* wird diesem hier mehr Platz eingeräumt. *Modellieren* ist in den Naturwissenschaftsdidaktiken noch nicht ausreichend einheitlich definiert, obwohl es in allen zu den zentralen Denk- und Arbeitsweisen gehört.

Im Gegensatz dazu wird *mathematisches Modellieren* weitgehend einheitlich als das „Aufstellen eines mathematischen Modells als geeignetes Abbild eines Ausschnitts der Welt“ bzw. als angewandtes „Problemlösen“ (Blum, 2006, S. 9; Niss, 2012, S. 50) hinreichend klar bezeichnet. Hierbei handelt es sich um einen *Modellierungskreislauf* (Blum et al., 2002, S. 230), der sechs Schritte umfasst (vgl. Abbildung 1). Ähnliche Klarheit könnte m. E. auch für die Chemie- und Biologiedidaktik hilfreich sein.

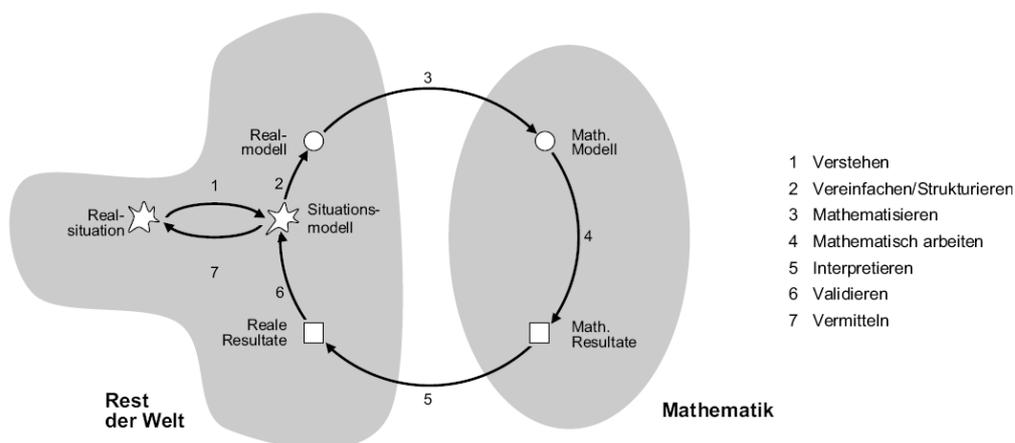


Abbildung 6: Modellierungskreislauf nach Blum (2006, S. 9).

Schritt 1 bildet die Konstruktion einer eigenen mentalen Vorstellung von der gegebenen Problemsituation mit dem Ziel, Situation und Fragestellung zu verstehen und hat ein Situationsmodell zur Folge. Schritt 2 beinhaltet, dass das resultierende *Situationsmodell* dann vom Schüler „strukturiert und vereinfacht“ wird. Hierdurch entsteht als Schritt 3 ein „*Realmodell* der Situation“ (Blum, 2006, S. 10), welches in Schritt 4 in ein *mathematisches Modell* über-

setzt wird. Hierbei „werden die Objekte und Annahmen mathematisiert“ (Blum, 2006, S. 10). In Schritt 5 wird mathematisches Wissen eingesetzt, um im Schritt 6 mathematische Resultate zu erhalten (Blum, 2006, S. 10; vgl. außerdem Niss, 2012, 49 f.).

Ähnliche Genauigkeit und Einigkeit ist bislang weder in der Chemie- noch in der Biologiedidaktik vorzufinden. Zwar schlägt Johnstone (1991) für die Chemiedidaktik einen Blum (2006) ähnelnden Modellierungskreislauf vor, der sich auch auf andere Naturwissenschaftsdidaktiken übertragen lässt, bislang aber weniger Beachtung findet als der Modellierungskreislauf von Blum (2006). Johnstone (1991) zielt auf Denkmodelle, darüber hinaus auf die Entstehung von wissenschaftlichem Denken bei Schülerinnen und Schülern generell. Hierbei wird aus einem für die Schülerinnen und Schüler neuen Unterrichtsinhalt (Phänomen, Idee, Experiment usw.) ein auf einem Denkmodell basierendes Präkonzept (eine selektive Vorstellung...), dessen wissenschaftlicher Inhalt nicht voll durchdrungen wurde, dessen Grad der Passung zur wissenschaftlichen Vorstellung jedoch durch Prozesse der Speicherung und des Bewusstwerdens zunimmt. Optimalerweise wird in der Schule spiralcurricular durch dort vermittelte Strategien, Konzepte, Fachsprache usw. sowie durch erneute selektive Durchdringung und bewusste Erarbeitung irgendwann eine wissenschaftlich adäquate Modellvorstellung erreicht.

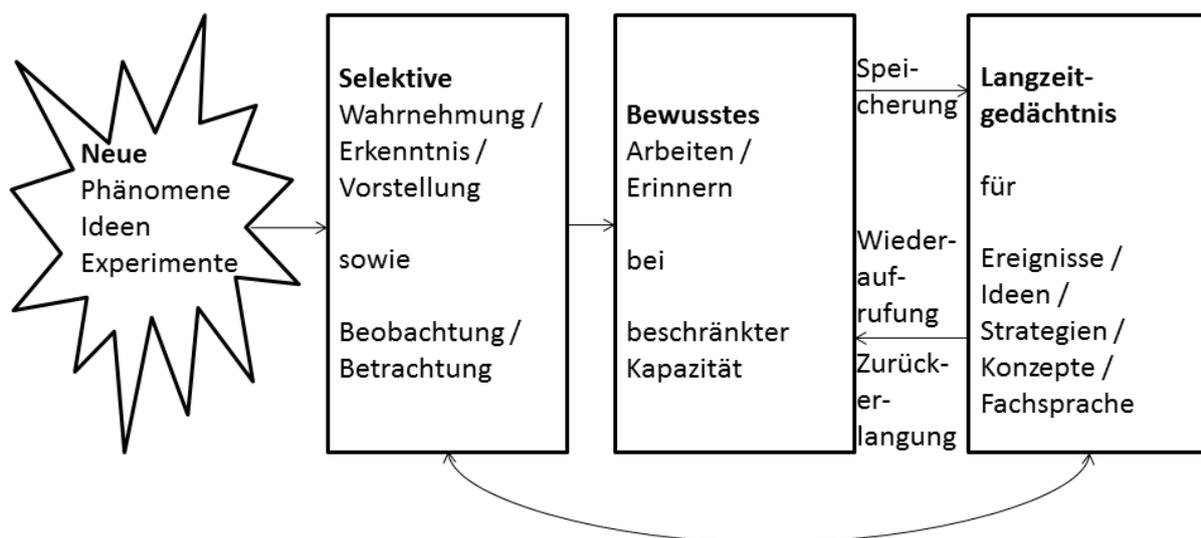


Abbildung 7: Modellierungskreislauf in der Chemiedidaktik nach Johnstone (1991, S. 81); adaptierte Übersetzung: B. T, eigene Darstellung.

3.3 Zusammenfassung und Schlussfolgerungen aus dem Theorieteil

Im Theorieteil dieser Studie konnte das Forschungsdesiderat eines handlungsnahen, empirisch-quantitativen und validen Lehrkompetenztests zur *Vermittlung von Modellkompetenz* dargestellt werden. Als Konstrukt wird hier das der Professionellen Unterrichtswahrnehmung, bestehend aus Reasoning und Noticing, gewählt. Professionalität zeichnet sich dabei als Anpassungsfähigkeit an Unterrichtsgegenstände (hier: *Vermittlung von Modellkompetenz*) aus. Daraus leitet sich die Notwendigkeit einer Noticing-Operationalisierung ab, die über ein systematisches Literaturreview erstellt wurde und die *Elemente der Vermittlung von Modellkompetenz* aufklärt. Diese Elemente dienen als Grundlage für die Entwicklung der Vignetten.

Im nun folgenden methodischen Teil der Arbeit und insbesondere im Ergebnisteil wird die Stimmigkeit der Theorie auf die erhobenen Daten empirisch-quantitativ überprüft. Hierzu wird das statistische Verfahren der konfirmatorischen Faktorenanalyse (CFA) als Teil der Strukturgleichungsmodellierung verwendet. Dabei werden die Stimmigkeit und Passung der empirischen Daten zum theoretischen Modell über die Fitindizes RMSEA (Root-Mean-Square Error of Approximation), CFI (Comparative Fit Index) und TLI (Tucker-Lewis Index) bestimmt. Laut Hu und Bentler (1998) geht der Fitindex RMSEA (Schwellenwert: $<.05$ (S. 449)) von einer bekannten Verteilung aus, gleicht Effekte hinsichtlich der Modellkomplexität aus und basiert dabei nicht auf Zentralität (S. 428). Der TLI (Hu & Bentler, 1998, S. 449, Schwellenwert: > 0.9) normiert nicht (also sind auch Werte $</> 0$ bzw. 1 möglich); zugleich gleicht er Effekte hinsichtlich Modellkomplexität aus (S. 428), wohingegen der CFI (Hu & Bentler, 1998, S. 449, Schwellenwert: > 0.9) normiert, also nur einen Range zwischen 0 und 1 vorweist und dabei nicht auf Zentralität basiert (S. 428). Laut Hu und Bentler (1998, 436 f.) korrelieren CFI und TLI trotz dieser Unterschiede in den repräsentativen Vergleichswerten mit 0.994 und 0.999.

Wenn im Folgenden zu Validierungszwecken von einem signifikanten Anstieg des Testscores die Rede ist, so sind basierend auf der klassischen Testtheorie p -Werte $< .05$, bei marginalsignifikanten Anstiegen p -Werte $< .1$ gemeint. Außerdem wird im Folgenden von Kompetenzzuwachs die Rede sein, obwohl Einigkeit darüber besteht, dass (Lehr-)Kompetenz nicht direkt erhoben werden kann (vgl. Weinert, 2001). Dennoch ist davon auszugehen, dass mit Professioneller Unterrichtswahrnehmung ein entscheidender Teil von Lehrkompetenz erhoben wird (Sherin, 2007).

3.4 Forschungsfragen und Hypothesen

Diese Studie untersucht folgende Forschungsfragen:

1. Finden sich in den Antworten der befragten Lehramtsstudierenden und Anwärterinnen und Anwärtern auf das Lehramt (im Folgenden der Einfachheit halber „Lehramtsstudierende“ genannt) der Biologie und Chemie Hinweise auf die Passung des Konstrukts Professionelle Unterrichtswahrnehmung zur Vermittlung von Modellkompetenz?
 - a. Stimmt die Struktur der Antworten der Lehramtsstudierenden mit der theoriebasierten Struktur des Noticing (*Elemente der Vermittlung von Modellkompetenz*) überein?
 - b. Stimmt die Struktur der Antworten der Lehramtsstudierenden mit der theoriebasierten Struktur des Reasonings (*aus Diagnosemöglichkeiten, Unterrichtsbewertungen und Handlungsalternativen auswählen können*) überein?
2. Lassen sich Hinweise auf die Validität des hier vorgestellten Vignettentests finden?
 - a. Nimmt die Kompetenz zur Professionellen Unterrichtswahrnehmung in Abhängigkeit von der Anzahl der Fachsemester zu?
 - b. Nimmt die Kompetenz zur Professionellen Unterrichtswahrnehmung in Abhängigkeit von der Anzahl der geleisteten Lehrstunden zu?
 - c. Findet sich unter den Antworten der Expertinnen und Experten ein eindeutiges Antwortmuster als Referenzwert?
3. Wie beeinflusst beim vorliegenden Vignettentest der Umgang mit fehlenden Werten das Ergebnis?

Aus den Forschungsfragen leiten sich folgende Hypothesen ab:

H1a-I) Die Passung der aus einem systematischen Literaturreview gewonnenen Struktur des Noticing (*Elemente zur Vermittlung von Modellkompetenz*) lässt sich im Biologietest belegen.

H1a-II) Die Passung der aus einem systematischen Literaturreview gewonnenen Struktur des Noticing (*Elemente zur Vermittlung von Modellkompetenz*) lässt sich im Chemietest belegen.

H1b-I) Die Passung der Reasoningbestandteile (*aus Diagnosemöglichkeiten, Handlungsmöglichkeiten und Unterrichtsbewertungen auswählen können*) lässt sich im Biologietest belegen.

H1b-II) Die Passung der Reasoningbestandteile (*aus Diagnosemöglichkeiten, Handlungsmöglichkeiten und Unterrichtsbewertungen auswählen können*) lässt sich im Chemietest belegen.

H2a-I) Der Testscore aus dem Biologietest steigt querschnittlich in Abhängigkeit von der Anzahl der Fachsemester an.

H2a-II) Der Testscore aus dem Chemietest steigt querschnittlich in Abhängigkeit von der Anzahl der Fachsemester an.

H2b-I) Der Testscore im Biologietest steigt querschnittlich in Abhängigkeit von der Anzahl der geleisteten Lehrstunden an.

H2b-II) Der Testscore im Chemietest steigt querschnittlich in Abhängigkeit von der Anzahl der geleisteten Lehrstunden an.

H2c-I) Die Antworten der Expertinnen und Experten der Biologiedidaktik bilden ein eindeutiges Antwortmuster als Referenzwert.

H2c-II) Die Antworten der Expertinnen und Experten der Chemiedidaktik bilden ein eindeutiges Antwortmuster als Referenzwert.

H3-I) Die Hauptaussagen der Ergebnisse aus dem Biologie- und Chemietest finden sich unabhängig vom Umgang mit fehlenden Werten wieder.

H3-II) Die Testschwierigkeit liegt unabhängig vom Umgang mit fehlenden Werten im mittleren Schwierigkeitsbereich.

Teil II: Methoden, empirisch-statistisches Vorgehen und Ergebnisse

4 Methoden

In diesem Kapitel werden die Vignettenentwicklung, die Entwicklung und der Einsatz von Kovariaten sowie die Validierungsschritte und die Bepunktung dargestellt.

4.1 Vignettenentwicklung

Bei der Entwicklung der Vignetten wurden sowohl das Noticing (*Elemente zur Vermittlung von Modellkompetenz*) als auch das Reasoning (*aus Diagnosemöglichkeiten, Handlungsmöglichkeiten und Unterrichtsbewertungen auswählen können*) berücksichtigt. Das Strukturdiagramm des Noticings diente als thematischer Hintergrund für den in den Vignetten beschriebenen Unterricht. Zur Entwicklung der Unterrichtshandlung wurde ein Ausschnitt aus einem realen, videografierten Unterricht (beispielsweise zur (Dis-)Kontinuumsproblematik) genutzt oder es wurden Hinweise aus der Literatur (z. B. zu Präkonzepten) eingebaut. Die Dimensionen des Reasoning wurden zur Entwicklung der Fragestellungen für die Itembatterien genutzt. Von einigen Studien wurde die Wirkung bestimmter Namen auf die Einstellung von Lehrkräften zu Schülerinnen und Schülern, die diese Namen tragen, und auf deren Schulerfolg untersucht (vgl. Kube, 2009). Daraus abgeleitet wurde versucht, bei den Probandinnen und Probanden diesbezügliche psychologische Mechanismen wie sich selbsterfüllende Prophezeiungen¹⁴ zu verringern. Aus diesem Grund wurden bei der Entwicklung der Vignetten die Vornamen der Schülerinnen und Schüler kontrolliert, d. h. die 23 beliebtesten Mädchen- und Jungennamen des Jahres 2000 wurden berücksichtigt (Bielefeld, 2016). Außerdem wurden etwa zehn bis 15 Prozent der Namen eingesetzt, die auf einen Migrationshintergrund schließen lassen. Dies unterrepräsentiert Menschen mit Migrationshintergrund, deren Anteil laut Mikrozensus (Statistische Ämter des Bundes und der Länder, 2011, S. 10) bei ca. 19,5 % der deutschen Bevölkerung liege, da über die Alterstruktur von Menschen mit Migrationshintergrund im Mikrozensus keine Aussagen getroffen wurden und der evtl. Effekt selbsterfüllender Prophezeiungen möglichst gering gehalten werden sollte. Die Ausführungen dieses Abschnitts gelten nur für die ursprünglich entwickelten Vignetten. Das Namens- und Zahlenverhältnis hat sich durch die Streichung von Vignetten im Rahmen der Validierung verändert.

¹⁴ Damit ist das in der Psychologie „Self-fulfilling prophecy“ genannte Phänomen gemeint.

Insgesamt wurden 51 Unterrichtsvignetten sowie kontextärmere, also weniger handlungsnah und ausführliche Darstellungen von Unterrichtssituationen, hinsichtlich Modellunterscheidung als Teil des Wissens um Nature of Science (NoS) mit 716 Items in einem geschlossenen Aufgabenformat entwickelt. Grund hierfür war das gezielte Adressieren der einzelnen *Elemente zur Vermittlung von Modellkompetenz*. Davon verblieben zur Haupterhebung 27 Unterrichtssituationen mit 217 Items in den Tests zur Biologiedidaktik (elf eigentliche Vignetten mit 115 Items sowie acht kontextärmere Unterrichtssituationen mit 33 Items zu Modellunterscheidung bzw. zu NoS) und im Test zur Chemiedidaktik acht Vignetten mit 69 Items. Im Durchschnitt gab es pro Unterrichtssituation also ca. acht Items, z. T. mit unterschiedlichen, aus dem Reasoningkonstrukt abgeleiteten Fragestellungen zu Itembatterien. Der Biologie- und der Chemietest enthalten sowohl Text- als auch Videovignetten. Im Biologietest ist der Videovignetten-Anteil geringer (drei Videovignetten von elf (27,3 %), unter Berücksichtigung der Modellunterscheidungsaufgaben drei von 19 Vignetten (15,8 %)) als im Chemietest (drei von acht Vignetten (37,5 %)), da im Biologieunterricht auch Paper-Pencil-Testhefte ohne Videos eingesetzt wurden. Videovignetten wurden nur in den Onlineumsetzungen (SoSci Survey: Leiner & Leiner, 2015) genutzt, die sowohl für den Biologie- als auch für den Chemietest entwickelt wurden.

Alle Textvignetten sind gleich aufgebaut (vgl. Abbildung 8 (Biologie) und Abbildung 9 (Chemie)). Videovignetten unterscheiden sich von den Textvignetten allein darin, dass statt des Vignettenstamms ein Video mit Kontext und Unterrichtsablauf gezeigt wird.

Killerzelle							}	Überschrift
Schülerinnen und Schüler einer neunten Klasse haben sich seit einigen Stunden mit dem Zusammenspiel verschiedener Zellen bei der Immunantwort beschäftigt. Die Klassenarbeit wird in einer Woche geschrieben. Zur Vorbereitung sollen die Schüler(innen) ein Begriffsnetz erstellen. Hierzu sollen die Schüler-innen und Schüler Fachbegriffe der Immunbiologie in Beziehung zueinander setzen, unter anderem auch die Begriffe „T-Killerzellen“ und „Grippe“. Danach sollen sich die Schülerinnen und Schüler ihr Begriffsnetz in Partnerarbeit gegenseitig erklären. Die Lehrerin tritt an den Tisch und hört zu, wie Lena zu Jonas sagt:								}
Lena: „Die T-Killerzellen killen Grippeviren.“							}	
Lehrerin: „Wie machen die T-Killerzellen das?“								}
Lena: „Na, wie wohl, die vernichten die Grippeviren.“							}	
<i>Diagnostizieren Sie Lenas Aussage. Machen Sie in jeder Zeile ein Kreuz.</i>								}
			Trifft überhaupt zu	überhaupt nicht	Trifft voll und ganz zu		}	
Lena verwechselt die Funktionen der Killer- und Fresszellen.			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
Lena übergeneralisiert die Nachahmungsfunktion der immunbiologischen Begriffe.			<input type="checkbox"/>					
Lena hat sprachlich auf die Funktion der Killerzelle geschlossen.			<input type="checkbox"/>					

Abbildung 8: Biologie-Textvignette zur Noticing-Facette *Präkonzepte* mit typischem Aufbau und Itemauswahl, eigene Darstellung.

Die folgende Abbildung 9 dient als Beispiel für eine Chemievignette und fasst deren Aufbau zusammen.

<u>Glucoseketten</u>		}		Überschrift	
Ein Lehrer lässt eine 10. Klasse, die sich bereits mit dem Aufbau von Stärkemolekülen und der Wirkungsweise von Enzymen beschäftigt hat, ungesäuertes und ungewürztes Brot mehrere Minuten lang kauen.		}		Kontext	
Lehrer: „Kann mir jemand erklären, warum das Brot am Ende süßer als am Anfang geschmeckt hat? (Wartet) Niemand? erinnert euch: Stärke besteht aus langen Ketten von Traubenzuckern.“		}		Vignettenstamm	
Jasmin: „Aber das ist doch unlogisch. Wenn wir Bonbonketten aus Traubenzucker auf dem Jahrmarkt kaufen, dann schmecken die doch sofort süß. Und nicht erst, wenn wir Minuten lang darauf herumkauen.“		}		Situation	
Lehrer: „Jasmin, das kann man so nicht vergleichen.“		}			
<i>Bewerten Sie den Unterricht aus fachdidaktischer Sicht. Machen Sie in jeder Zeile ein Kreuz.</i>		}		Frage	
		Trifft überhaupt nicht zu	Trifft voll und ganz zu	sechsstufige Skala	
„Kette“ ist ein verwirrendes Wort und sollte aus didaktischen Gründen hier vermieden werden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Der Versuch sollte vor der Theorie des Stärkeaufbaus stehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Mit der Aussage „Stärke besteht doch aus langen Ketten von Traubenzuckern“ bleibt der Lehrer auf der makroskopischen Ebene.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Items

Abbildung 9: Chemie-Textvignette zur Noticing-Facette *Präkonzepte* mit typischem Aufbau und Itemauswahl, eigene Darstellung.

Es ist nicht möglich, alle entwickelten Vignetten und Items in dieser Dissertationsschrift abzubilden, da es sich um einen Leistungstest handelt, der weiter im Einsatz sein wird.

4.2 Kovariaten

An dieser Stelle wird die Gestaltung, Zielrichtung und der Grund für die Entwicklung des Wissenstests über Modellunterscheidung als Teil des Wissens über die Natur der Naturwissenschaft (Nature of Science, NoS) im Biologieunterricht erklärt. Weitere Themen dieses Kapitels sind die Begründung, die Beschreibung und Annahmen beim Einsatz der Kovariaten.

Es gibt Evidenz dafür, dass Wissen über Modelle Einfluss nimmt auf die Fähigkeit, Modellkompetenz an Schülerinnen und Schülern zu vermitteln (Henze et al., 2007, 2008; Jong & van Driel, 2004; Jong, van Driel & Verloop, 2005a; Justi & van Driel, 2005a, 2005b, 2005c; Riehs, 2013). Die genannten Artikel stammen allesamt aus der Chemiedidaktik. In der Biologiedidaktik finden sich weniger und indirekter Forschungsergebnisse zu diesem Zusammenhang (vgl. z. B. Schmelzing et al., 2013a), obwohl beispielsweise auf die Bedeutung von Modellunterschieden hingewiesen wurde (Meisert, 2008, 2014). Es fehlt jedoch ein Test, der be-

zogen auf das biologische Fachwissen für den Biologieunterricht Modellunterscheidung in den Fokus nimmt. Deswegen wurde solch ein Test für den Biologieunterricht entwickelt. Dieser nimmt das Wissen über Modellunterscheidung (vgl. Abbildung 10) in den Fokus.

Mitochondrien-Kraftwerk

Ein Lehrer trifft in einer 8. Klasse folgende Aussage: „Mitochondrien sind wie kleine Kraftwerke, die Energie für die Zelle produzieren.“

Handelt es sich bei dieser Aussage aus fachdidaktischer Sicht um...

	Trifft gar nicht zu	Trifft voll zu
ein Homologmodell?	<input type="radio"/>	<input type="radio"/>
ein Analogmodell?	<input type="radio"/>	<input type="radio"/>
ein Strukturmodell?	<input type="radio"/>	<input type="radio"/>

Abbildung 10: Beispiel eines Teils einer Aufgabe aus dem Wissenstests über Modellunterscheidung im Kontext des Wissens über die Natur der Naturwissenschaften (Nature of Science, NoS) im Biologieunterricht, umgesetzt in Sosci-Survey (Leiner & Leiner, 2015).

Aufgaben zu Modellunterscheidungen wie in Abbildung 10 dargestellt, wurden ergänzt um Einschätzungen wie beispielsweise zu Analogien und der Unterscheidung von Realität und Modellen.¹⁵

Grund für die Nichtentwicklung eines Chemietests zum *Wissen über Modellunterscheidungen* war neben der bereits vorhandenen Evidenz ein zeitökonomischer: Weil es verhältnismäßig wenige Chemielehramtsstudierende gibt, sollte über eine kurze Testzeit möglichst sichergestellt werden, dass ein hoher Anteil der Befragten den Chemietest vollständig ausfüllt.

Die ebenfalls aus Validitätsgründen eingesetzten Kovariaten wurden als Ganzes oder in Teilen übernommen. Im Folgenden wird der Grund für den Einsatz der Kovariaten kurz umrissen und es werden Hypothesen einer erwarteten Korrelation des Testscores mit den Kovariaten genannt.

Mindestens seit Shulman (1987, S. 8) fachdidaktisches Wissen bzw. PCK als „special amalgam“ (vgl. Kapitel 2.1.1, S. 19) bezeichnete, ist eine Diskussion über die „fuzzy bounda-

¹⁵ Vgl. hierzu Kapitel 3.2.3 F, S. 52

ries¹⁶ (Gess-Newsome, 1999, S. 10) von PCK zwischen Fachwissen (CK) und pädagogisch-psychologischem Wissen (PK) im Gange. Insofern sind auch in dieser Studie Tests sowohl zu CK als auch zu PK als Kovariate eingesetzt worden. Einen passenden CK-Test für das Schulfach Biologie zu bekommen, war nicht einfach. Geeignete Instrumente waren noch nicht veröffentlicht und konnten deshalb nicht genutzt werden (vgl. z. B. Jüttner, 2012). Deshalb wurde aus bereits vorhandenen Items (ETS, 2015a) über mehrere Pilotierungsschritte ein CK-Test mit einer breit angelegten Themenauswahl entwickelt. Erwartet werden im Gegensatz zu Largescale-Studien kein hoher, sondern ein moderater Zusammenhang zwischen CK-Test und Biologietestscore sowie ein etwas höherer Zusammenhang zwischen CK-Test und NoS-Wissen zur Modellunterscheidung. Grund dafür ist, dass wir für unterrichtsnahe Vignettentests von einer geringeren Korrelation zum Fachwissen ausgehen als dies bei nah am Fachwissen operationalisierten fachdidaktischen Largescale-Studien der Fall ist. Außerdem sollte NoS-Wissen näher mit CK zusammenhängen. Für den Chemietest konnten wir auf den CK-Test von Witner und Tepner (2011c) zurückgreifen, ergänzt um drei Items von ETS (2015b), um eine höhere Curriculum-Passung zu erreichen.

Ebenfalls um einen Beitrag zur Aufklärung der verschwommenen Grenzen von PCK (Gess-Newsome, 1999, S. 10) zu leisten, wurde ein PK-Test gesucht. Um die Testzeit möglichst gering zu halten, sollte dieser nur geschlossene Items enthalten. Da der bei Voss, Kunter und Baumert (2011) beschriebene PK-Test nicht zu bekommen war, wurden geschlossene Items des Tests „Pädagogisches Unterrichtswissen“ (PUW) (König & Blömeke, 2010) eingesetzt; die offenen Items wurden aus Testzeitgründen entfernt. Es wird ein schwacher bis moderater Zusammenhang zwischen dem PK-Test und dem Biologie- und Chemietest angenommen.

Um auszuschließen, dass der entwickelte Vignettentest letztlich über Intelligenz und im Speziellen bei Textvignetten über sprachliche Fähigkeiten gelöst werden könnte, ist es notwendig zu zeigen, dass zur Lösung des Vignettentests eine für sich stehende Fähigkeit und damit ein eigenständiges Konstrukt erhoben wird. Deswegen wurde ein Testinstrument zur Erhebung von Intelligenz in Verbindung mit sprachlichen Fähigkeiten gesucht und mit dem „Wilde-Intelligenz-Test“ (Kersting, Althoff & Jäger, 2008), der auf dem Erkennen von Wortanalogien beruht, gefunden. Erwartet wird ein geringer, aber signifikanter Zusammenhang, der evtl. auch moderiert wird über den angenommenen engeren Zusammenhang zwischen Vignettenscores und Fachwissen.

¹⁶ Verschwommene Grenzen von PCK (Übersetzung B. T.)

Da als Voraussetzung zum Wissen über Modellunterscheidungen und im Speziellen über Vermittlung von Modellkompetenz ein hohes Maß an Fähigkeit zum abstrakten Denken angenommen werden kann, welches mit der Freude am Denken einhergehen könnte, wird eine schwache bis moderate Korrelation zwischen dem Kognitionsbedürfnis (Cacioppo & Petty, 1982) der künftigen Lehrkräfte und den Vignettentestscores vermutet. Ebenso könnten bei der Vermittlung von Modellkompetenz das Selbstkonzept, die Motivation und die Persönlichkeit der künftigen Lehrkräfte eine Rolle spielen. Es werden eher schwache als moderate Korrelationen zum Berufsselbstkonzept (Retelsdorf, Bauer, Gebauer, Kauper & Möller, 2014) und zur Motivation der Aufnahme eines Lehramtsstudiums (Pohlmann & Möller, 2010) erwartet bzw. zu einzelnen Subskalen dieser Kovariaten. Gleiches gilt für die Persönlichkeitsaspekte (vgl. Gerlitz & Schupp, 2005). Denn dass solche Aspekte Einfluss nehmen können, ist anzunehmen; ein enger Zusammenhang würde jedoch der Hypothese der Eigenständigkeit des mit Vignettentests erhobenen Konstrukts widersprechen.

4.3 Validierungsschritte

Da bei der Testentwicklung Validitätsfragen im Vordergrund standen, werden in diesem Kapitel die Validierungsschritte chronologisch entsprechend der Vignettenentwicklung dargestellt. Die Frage der Validität, also „wie gut der Test in der Lage ist, genau das zu messen, was er zu messen vorgibt“ (Bortz & Döring, 2006, S. 200), stellt sich bei Vignettentests in besonderer Weise. Zu beachten sind wie bei anderen Tests auch die statistische Rückschluss-Validität¹⁷, die interne Validität¹⁸, die Konstruktvalidität¹⁹, die Populationsgeneralisierbarkeit²⁰, die zeitliche Generalisierbarkeit²¹, die Aufgabengeneralisierbarkeit²² sowie die externe Validität²³ und die ökologische Generalisierbarkeit²⁴. Die beiden letztgenannten Maße werden

¹⁷ Vgl. Messick (1989, S. 57) zur statistischen Rückschluss-Validität: Ist das statistische Ergebnis unabhängig von möglichen Fehlern bei der Stichprobenziehung?

¹⁸ Vgl. Messick (1989, S. 57) zur internen Validität: Die statistischen Unterschiede zwischen Teilstichproben sind nur durch die erwarteten bzw. plausiblen Effekte verursacht.

¹⁹ Vgl. Messick (1989, S. 58) zur Konstruktvalidität: Handelt es sich um generalisierbare Rückschlüsse von der Messung auf Ursache-Wirkung-Zusammenhängen?

²⁰ Vgl. Messick (1989, S. 56) zur Populationsgeneralisierbarkeit: Das empirisch gemessene Konstrukt lässt sich in einer anderen Population ebenfalls wiederfinden.

²¹ Vgl. Messick (1989, S. 56) zur zeitlichen Generalisierbarkeit: Die Messergebnisse lassen sich in einer weiteren, zeitversetzten Erhebung wiederholen.

²² Vgl. Messick (1989, S. 56) zur Aufgabengeneralisierbarkeit: Mit Aufgaben aus einem anderen Test lassen sich vergleichbare Ergebnisse wiederfinden.

²³ Vgl. Messick (1989, S. 58) zur externen Validität: Wie generalisierbar ist der Zusammenhang von Populationsgruppen, Erhebungsumständen und Erhebungszeitpunkten? Hier: Korrespondieren die Vignetten mit zentralen und realen Unterrichtssituationen?

zwar durch das Vignettenformat begünstigt (Rehm & Bölsterli, 2014, S. 213), sind jedoch nicht per se ausreichend vorhanden.

Am Anfang stand daher die Sichtung von Unterrichtsvideos zum Modelleinsatz. Entscheidende Szenen, die als prototypisch gelten können, wurden ebenso als Grundlage für Ideen zur Vignettenentwicklung genutzt wie Hinweise, die aus der Literaturrecherche kamen.

Es folgte eine qualitative Expertinnen- und Expertenbefragung hinsichtlich fachdidaktischer Relevanz sowie zur Nähe zum Unterrichtsalltag. Expertise wurde in diesem Schritt neun Lehrbeauftragten der Staatlichen Seminare für Lehrerbildung Karlsruhe und Ludwigsburg zugesprochen. Sie und eine Expertin und ein Experte für Modelle im Biologie- und Chemieunterricht an einer schweizer Hochschule wurden befragt, um die Anschlussfähigkeit im deutschsprachigen Raum zu ermöglichen. Ziel war es, insbesondere dem Anspruch auf ökologische Generalisierbarkeit zu entsprechen und mit den Vignetten an die Unterrichtspraxis anzuschließen. In der Folge wurden die Unterrichtssituationen entsprechend den Ausführungen der Befragten angepasst und Items verändert bzw. neu formuliert.

Im nächsten Validierungsschritt (Populationsgeneralisierbarkeit) wurde eine quantitative Expertinnen- und Expertenbefragung über das Onlinebefragungswerkzeug SoSci Survey (Leiner & Leiner, 2015) durchgeführt. Um dem Anspruch auf Populationsgeneralisierbarkeit gerecht zu werden und eine breite Sicht auf Modelle im Biologie- und Chemieunterricht zu gewährleisten, wurde nun die Expertise neben den Fachdidaktikerinnen und Fachdidaktikern aus Universitäten und Hochschulen in Deutschland, der Schweiz und Österreich auch Lehrbeauftragten Staatlicher Seminare für Lehrerbildung (Referendariatsausbildung) und Lehrkräften (Bromme, 2004) zugeschrieben. Insgesamt wurden ca. $N = 450$ Expertinnen und Experten angefragt; 45 Chemiedidaktik- und 60 Biologiedidaktik-Expertinnen und -Experten nahmen an der Befragung teil.

Nach der Beantwortung auf einer sechsstufigen Skala zwischen „Trifft gar nicht zu“ und „Trifft voll zu“ diente der Modalwert auf Itemebene als erster Referenzwert für den Vignettest Biologie und Chemie. Entsprechend wurden Items bei uneindeutigem Modalwert entfernt. Uneindeutigkeit wurde zum einen über die Standardabweichung vom Mittelwert (> 1.5) sowie über die Trennschärfe ($< .3$) bestimmt. Drittes Selektionskriterium war die Abweichung vom Modalwert: Lag der Betrag der Abweichung > 2 bei $\geq 25\%$, so wurde das Item ebenfalls entfernt (Tepner & Dollny, 2014; vgl. Witner & Tepner, 2011b). Außerdem waren auch quali-

²⁴ Vgl. Messick (1989, S. 56) zur ökologischen Generalisierbarkeit: Die Übertragbarkeit der Messung auf andere Situationen, hier die Fähigkeit zum adäquaten Unterrichten von Modellen.

tative Rückmeldungen möglich, welche in den Test eingearbeitet wurden. Der Test wurde auf der Grundlage des Antwortverhaltens von $N = 38$ Biologielehramtsstudierenden pilotiert. Dadurch konnte Verständnisproblemen begegnet werden. Außerdem wurden dadurch Rückschlüsse auf die Auswertungsmöglichkeiten gezogen. Zum Beispiel wurde die Bepunktung der Abweichung vom Referenzwert über ein Partial-Credit-Verfahren (0/0,5/1 Punkt) für jede Vignette (Gesamtsummenscore 92 Punkte, $M = 43,6$; $SD = 5,77$; Range: Min = 19,5, Max = 56) erprobt.

Ein weiterer Validierungsschritt war eine zweite quantitative Expertinnen- und Expertenbefragung. Nach grundlegender Überarbeitung des Vignettentests (u. a. Ergänzung von Video-vignetten) sowohl in den Fachdidaktiken Biologie als auch in Chemie wurden nun besonders ausgewiesene Expertinnen und Experten (gemessen an den Publikationen zum Modelleinsatz) befragt. Die ausgewählten Expertinnen und Experten wurden wiederum durch einen Link des Onlinetools SoSci Survey eingeladen. Sie beantworteten die Items und gaben qualitative Rückmeldungen. Aus den Antworten ($N_{\text{Biologie}} = 9$; $N_{\text{Chemie}} = 21$) wurde die endgültige Expertinnen- und Expertennorm gewonnen. Das Vorgehen bei der Auswertung entspricht dem, welches bzgl. der ersten quantitativen Befragung dargestellt wurde (s. o.).

Auf der Grundlage der Expertinnen- und Expertenbefragung wurden unzureichende Vignetten und Items entfernt. Als erstes Selektionskriterium der Items wurden die Standardabweichung vom Mittelwert > 1.5 sowie die Trennschärfe ($< .3$) herangezogen. Zweites Selektionskriterium war die Abweichung vom Modalwert: Lag der Betrag der Abweichung > 2 bei $\geq 25\%$, so wurde das Item ebenfalls entfernt (vgl. Tepner & Dollny, 2014; Witner & Tepner, 2011c). Auf dieser Grundlage wurde beispielsweise Item E9V24I03 nicht weiter bei der Auswertung berücksichtigt, da die befragten Expertinnen und Experten kein eindeutiges Antwortmuster aufwiesen, während E9V23I05 in der Auswertung berücksichtigt wurde (vgl. Abbildung 11).

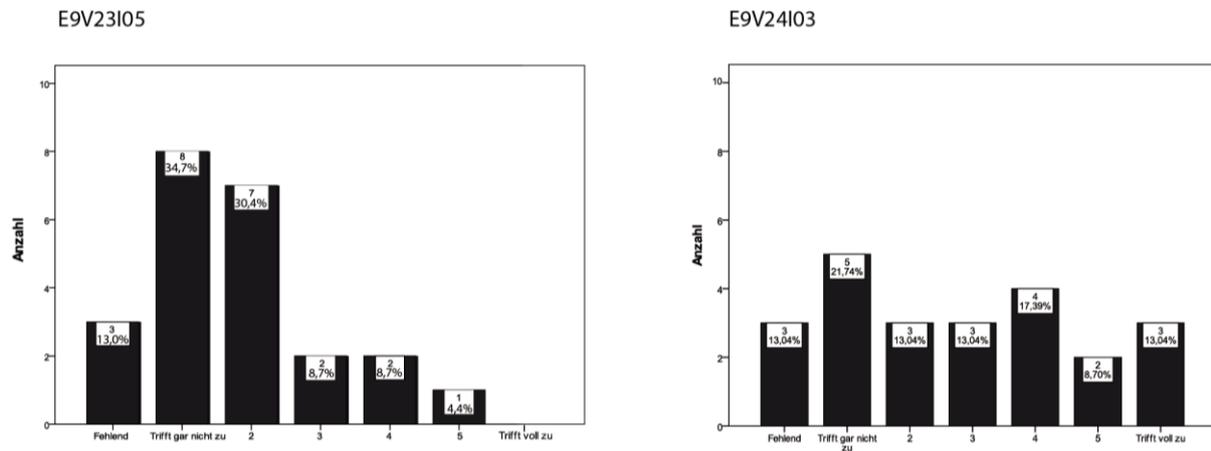


Abbildung 11: Antwortverteilung der Expertinnen und Experten: Aufgrund eindeutiger Antworten konnte das Item E9V23I05 bei der weiteren Auswertung berücksichtigt werden, Item E9V24I03 hingegen nicht, eigene Darstellung.

4.4 Bepunktung

Die „Bepunktung“ (Tepner & Dollny, 2014; Witner & Tepner, 2011c) eines Vignettentests ist komplex, da es bei der Bewertung von Unterrichtssituationen zunächst keine objektive Aussage darüber gibt, welche Antwort richtig oder falsch ist. Aus diesem Grund wurde eine sechsstufige Skala zwischen „trifft gar nicht zu“ und „trifft voll und ganz zu“ genutzt. Expertinnen und Experten wurden gebeten, den Vignettentest zu beantworten, um einen Referenzwert bilden zu können. Hierbei wird davon ausgegangen, dass ähnlich wie bei der Demokratie die Mehrheit der Expertinnen und Experten eher Aussagen über die Korrektheit einer Antwort geben kann als einzelne Expertinnen und Experten.

Die von den Expertinnen und Experten ausreichend eindeutig beantworteten Items wurden weiter berücksichtigt. Aus den aggregierten Expertinnen- und Expertenurteilen wurden im zweiten Schritt Modalwerte berechnet. Nach der hier vertretenen Meinung ist nicht der absolute Modalwert, sondern das Verhältnis zweier Items zueinander entscheidend. Dies greift die Erfahrung auf, dass Lehrkräfte während des Unterrichts zwischen verschiedenen Möglichkeiten der Aktion oder Reaktion auswählen müssen. Entspricht das Verhältnis zweier Items im Antwortverhalten der Probandinnen und Probanden (vgl. Abbildung 12) dem Verhältnis der Modalwerte der Expertinnen und Experten, wird ein Punkt vergeben (vgl. Abbildung 13).

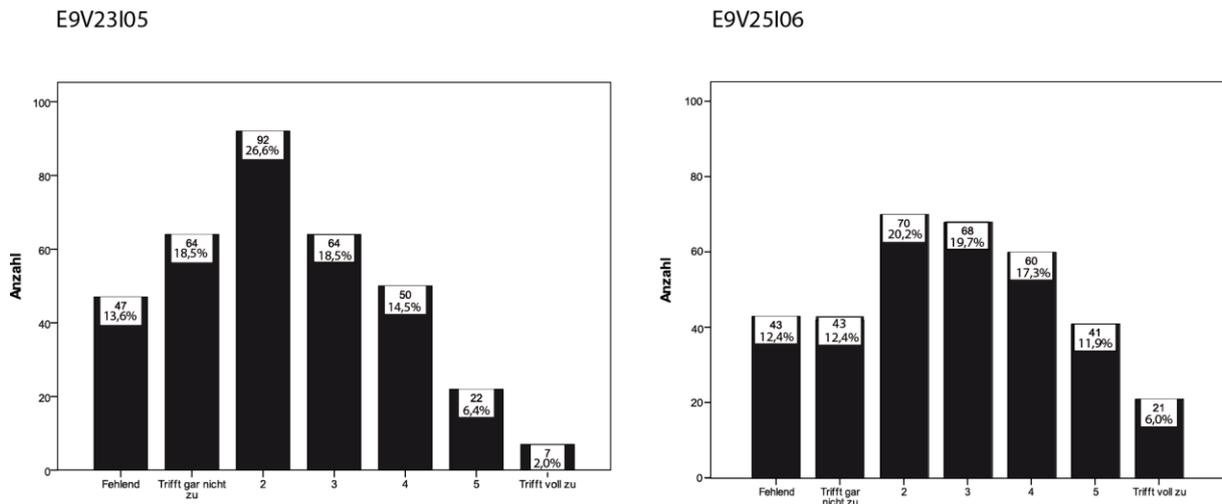


Abbildung 12: Vergleich zweier Items hinsichtlich der Verteilung der Antworten der Probandinnen und Probanden, eigene Darstellung.

Diese Bepunktungsmethode wird im Folgenden als Paarvergleich im Sinne einer Relation zweier Items zueinander bezeichnet (vgl. Artelt, Beinicke, Schlagmüller & Schneider, 2009, S. 98; Tepner & Dollny, 2014, S. 318; Thillmann, 2008, S. 94). Da nicht von einer zu großen Sensitivität ausgegangen werden kann (Witner & Tepner, 2011a, S. 126), wurden Punkte nur bei Modalwertunterschieden > 1 gewertet (vgl. Abbildung 13).

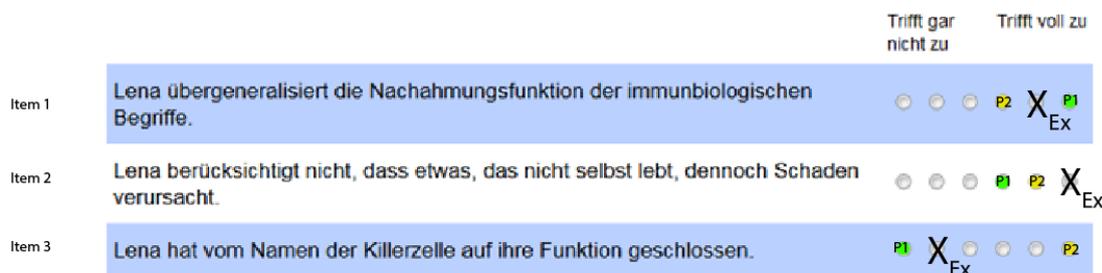


Abbildung 13: Hypothetisches Beispiel für die Bepunktung des Paarvergleichs (PV): Die Modalwerte des Expertinnen- und Expertenratings (X_{Ex}) werden je Itempaar miteinander verglichen. Proband 1 (P1) hat im Gegensatz zum aggregierten Expertinnen- und Expertenurteil nicht erkannt, dass die in Item 2 getroffene Aussage besser ist als die in Item 1 getroffene Aussage. Da sich die Modalwerte des Expertinnen- und Expertenratings jedoch nur im Betrag 1 voneinander unterscheiden, wird diese Relation nicht berücksichtigt. Entscheidend sind nur der PV zwischen Item 1 und Item 3 sowie Item 2 und Item 3. Proband 1 hat erkannt, dass Item 1 und Item 2 besser als Item 3 sind und bekommt hierfür jeweils einen Punkt. Im Gegensatz dazu erhält Proband 2 (P2) für beide Relationen je 0 Punkte, da er die in Item 3 getroffene Aussage für die beste Lösung hält.

Eine andere Möglichkeit der Bepunktung wäre beispielsweise ein Partial-Credit-Verfahren, wie es bei der ersten Befragungsrunde eingesetzt wurde (vgl. 4.3, S. 65). Hierbei bekamen Probandinnen und Probanden, die das Item gleich einschätzten wie der Modalwert der Exper-

tinnen und Experten einen Punkt. Probandinnen und Probanden mit einer Antwort, die im Betrag nur einen Punkt vom Modalwert abwich, erhielten einen halben und Probandinnen und Probanden mit größerem Abstand 0 Punkte. Die Entscheidung für den Paarvergleich liegt darin begründet, dass über diesen eine Nähe zur Unterrichtsrealität hergestellt werden kann. Im Unterricht steht eine Lehrkraft stets unter dem Druck der Abwägung zwischen verschiedenen Reaktionsmöglichkeiten. Beim Paarvergleich wird dies dadurch simuliert, dass keine absolute Richtigkeit, sondern nur das Abwägen zweier Items als besser oder schlechter erfolgt.

Die Paarvergleiche wurden auf der Ebene der Fragestellungen (vgl. Kapitel 3.2.2, S 40) summiert (im Folgenden „Paarvergleichssummen“ genannt). Dies geschah, um eine der theoretischen Voraussetzungen für Tests und im Speziellen für IRT zu erfüllen: die lokalstochastische Unabhängigkeit. Damit ist die „[l]ogische Unabhängigkeit zwischen den Items“ (Rost, 2004, S. 74) gemeint. Durch die präsentierten Unterrichtssituationen sowie deren verschiedene Fragestellungen zu Itembatterien wird eine logische Abhängigkeit der Items und Paarvergleiche wahrscheinlich. Zum Ausgleich werden Paarvergleichssummen gebildet, d. h. alle Paarvergleiche einer Fragestellung werden summiert. So wurden bspw. die Punkte der in Abbildung 13 dargestellten drei Items für alle Teilnehmenden summiert und zur Paarvergleichssumme PVS25B zusammengefasst. Mit diesem Pseudoitem wurde dann weitergerechnet. Sowohl im Chemie- als auch im Biologietest gab es jeweils eine Paarvergleichssumme, die mehr als zehn Kategorien enthielt. Da das Programm Mplus (vgl. Muthén & Muthén, 2012) in diesem Fall die probabalistische Berechnung unterbindet, wurden Kategorien gruppiert, d. h. beispielsweise wurden Fälle mit 0, 1, 2 und 3 Punkten in der Paarvergleichssumme PVS25B zu einer Kategorie zusammengefasst.

5 Statistisches Vorgehen

In diesem Kapitel werden die Stichprobe, die Analyseverfahren und der Umgang mit fehlenden Werten dargestellt.

5.1 Stichprobenbeschreibung

Im Biologietest wurden $N = 376$ Lehramtsstudierende für die Datenauswertung berücksichtigt. Je nach Umgang mit fehlenden Werten werden diese zur Datenauswertung herangezogen oder nur $N = 348$ (92,6 %), wenn als Voraussetzung das Ausfüllen von 50 % der Vignetten gewertet wird (vgl. Kapitel 5.3, S. 72). Als Ausfüllen einer Vignette wird das Beantworten mindestens eines Items gewertet, da so sichergestellt wird, dass keine technischen Gründe für das Nichtausfüllen sprachen. Im Chemietest wurden ursprünglich $N = 315$ Personen für die Datenauswertung berücksichtigt und später $N = 31$ Personen nacherhoben. So ergibt sich eine Gesamtstichprobe für den Chemietest von $N = 346$. Werden hiervon nur Personen, die 50 % der Vignetten ausfüllten, gewertet, verbleiben insgesamt $N = 310$ im Test (89,6 %). Die Gründe für die höhere Ausfallquote im Chemietest könnte in der ausschließlichen Beantwortung über das Internet gesehen werden (vgl. Kapitel 5.3 „Umgang mit fehlenden Werten“, S. 72). Die Ausfallquote ist im Vergleich zu anderen im Internet zu beantwortenden Vignetentests (Kaiser, 2016 berichtet von bis zu 21 Prozent Ausfall bei geschlossenen Items) moderat.

Tabelle 4 zeigt eine Übersicht über den Einfluss einiger zentraler soziodemografischer Aspekte auf die Ergebnisse des Chemie- und Biologietests. Der höhere Anteil von Frauen wird im Fach Biologie häufiger berichtet (vgl. z. B. Jüttner & Neuhaus, 2013, S. 39).

Tabelle 4: Zentrale soziodemografische Werte des Chemie- und Biologietests

Test	N	Alter		Geschlecht weiblich	Fachsemester	
		M	SD		M	SD
Chemie	310	24.05	4.27	58.4 %	5.04	2.39
Biologie	348	23.63	3.46	72.1 %	4.44	2.91

Anmerkung: Als Grundlage wurde jeweils ein Datensatz genutzt, bei dem Befragten mit weniger als 50 % fehlender Werte für jeden verbleibenden fehlenden Wert der Wert 0 imputiert wurde (vgl. Kapitel 5.3, S. 72).

5.2 Statistische Analyseverfahren und Software

Zur Berechnung der Validitätsaspekte wurde das Programm SPSS 23 (IBM Corporation, 2015) auf Grundlage der klassischen Testtheorie genutzt. Für die Überprüfung der Modellpas-

sung zu den empirischen Daten (Konstruktvalidität) wurde das Programm Mplus 7.3 (vgl. Muthén & Muthén, 2012) verwendet. Diese Berechnungen erfolgten probabalistisch (categorical und Estimator = WLSMV). Zur Berechnung der Test- und Itemschwierigkeit wurde das R-Package mirt (Chalmers, 2014) genutzt.

5.3 Umgang mit fehlenden Werten

„Während die Selektion von Items als legitim gilt, schließlich sind sie von Menschenhand gemacht und können mit allen Fehlern behaftet sein, die eine Eliminierung rechtfertigen, gilt die Eliminierung unpassender Personen aus der Datenmatrix als illegitim. Es liegt der Argwohn der Datenmanipulation nahe, wenn man von einer Stichprobe von Beobachtungen, das sind in diesem Fall die Testprotokolle, einfach einem [sic!] Teil wegläßt, um die Ergebnisse zu verschönern.“ (Rost, 1996, S. 381) Im Folgenden wird deshalb nach einem legitimen Umgang mit fehlenden Werten der Studie gesucht.

In der hier beschriebenen Studie wurden alle leeren Fälle (Probandinnen und Probanden, die nichts ausgefüllt hatten) entfernt. Vergleichsweise hoch war die Anzahl der Probandinnen und Probanden, die nur (Teile der) Kovariaten, nicht jedoch die Vignetten (Missingquote in Chemie: 7,9 %, in Biologie: 4,5 %) ausfüllten. Auch mit diesen fehlenden Werten muss im Sinne Rosts (1996) korrekt umgegangen werden.

Hierzu ist Information über den Testaufbau erforderlich. Die Kovariaten wurden, bis auf den Wortanalogientest (Gerlitz & Schupp, 2005), zeitlich nach dem Vignettentest ausgefüllt. Aufgrund der Rotation der Vignettentests bei der Onlineumsetzung (sowohl die Reihenfolge der Vignetten als auch der Items rotierte), die bei Biologie 60,1 % und bei Chemie 100 % ausmachte, ist dort ein „Missing Data by Design“ auszuschließen. Um weiter zu differenzieren, wurden fehlende Werte, bei denen mindestens ein Item einer Vignette ausgefüllt wurde, mit dem Wert -9 kodiert. War in einer Vignette kein Item ausgefüllt, wurde für den fehlenden Wert -8 vergeben. Dies geschah, um bei der Codierung mit -9 technische Probleme auszuschließen.

Insgesamt liegen die fehlenden Werte im Bereich Chemie zwischen 9,8 % und 15,0 % je Item, im Bereich Biologie bei 6,6 % bis 13,0 % (beim Modellunterscheidungstest – Wissensabfrage ohne Vignetten im Bereich NoS (vgl. Kapitel 3.2.3 F, S. 52) – gab es ein Ausreißeritem mit 21,5 %). Welche Gründe sind für das Auftreten fehlender Werte vorstellbar?

Formal gesehen wird in der Statistik zwischen drei verschiedenen Typen fehlender Werte unterschieden (Garson, 2015; Lüdtke, Robitzsch, Trautwein & Köller, 2007): Missing Completely At Random (MCAR), Missing At Random (MAR) und Missing Not At Random (MNAR). Wenn fehlende Werte als MCAR gelten können, ist mit diesen Werten am besten umzugehen. Denn das bedeutet, dass das Auftreten fehlender Werte zufällig, also normalverteilt ist. Dies kann über einen nicht signifikanten MCAR-Test (Little, 1988) überprüft werden. Allerdings werden beim MCAR-Test zwei Voraussetzungen im engeren Sinne nicht überprüft: Zum einen sollten die Testscores nicht hoch mit miterhobenen Kovariaten korrelieren (MAR), zum zweiten sollten auch keine hohen Korrelationen zu nicht miterhobenen Konstrukten (Kovariaten oder Soziodemografie) bestehen (MNAR).

Wenn fehlende Werte als MAR gelten, sollten folgende drei Voraussetzungen gelten. Erstens sollte der MCAR-Test (Little, 1988) signifikant werden, das heißt, für MCAR gilt die Annahme der Nullhypothese. Zweitens sollten die fehlenden Werte über miterhobene Kovariaten oder soziodemografische Angaben vorausgesagt werden können. Beispielsweise berichten Voss et al. (2011, S. 963) von einer hohen latenten Korrelation zwischen mathematischem Fachwissen und fachdidaktischem Wissen (.91). Fehlende Werte im fachdidaktischen Wissenstest könnten so relativ zuverlässig über die Werte im Fachwissenstest geschätzt werden. Drittens dürfen die fehlenden Werte nicht von anderen Variablen, die nicht miterhoben wurden, beeinflusst sein (MNAR).

Mit MNAR kann man am schwersten umgehen, weil man nicht um das Konstrukt weiß, welches auf die fehlenden Werte Einfluss nahm (sonst hätte man dieses nach den Gepflogenheiten der Testkonstruktion miterhoben). MNAR gilt also, wenn fehlende Werte von nicht miterhobenen Variablen beeinflusst werden und die miterhobenen Kovariaten für die Voraussage des Tests nicht herangezogen werden können. Hierdurch werden Möglichkeiten zur Imputation a priori unterhöhlt.

Ob es sich bei den hier berichteten fehlenden Werten um MCAR, MAR oder MNAR handelt, ist nur schwer zu bestimmen. Testspezifisch können folgende Gründe konstatiert werden: Die fallweise Häufung von fehlenden Werten könnte darin begründet sein, dass das Bearbeiten im Internet zu Hause zu einer geringeren Verbindlichkeit geführt hat. Im Biologietest lassen sich post-hoc höchst signifikante Unterschiede zwischen komplett fehlenden Werten bei Vignetten (-8) und den Testformaten Paper-Pencil und Computer zu Hause nachweisen. Dies spricht für die These des Einflusses der geringeren Verbindlichkeit zu Hause. Allerdings sollte dann auch ein signifikanter Unterschied zwischen dem Bearbeiten am Computer zu Hause und am Computer im Rahmen einer Seminarveranstaltung feststellbar sein. Dieser Unterschied ist im

Chemietest erwartungskonform hoch signifikant, jedoch im Biologietest nicht nachweisbar. Auch technische Probleme können insgesamt nicht ausgeschlossen werden.

Für die relativ hohe Anzahl fehlender Werte könnte außerdem auch das Format Vignettentest verantwortlich sein. So berichtet Kaiser (2016) beim Videovignettentest der Studie TEDS-FU (Teacher Education and Development Study/Follow Up) bei geschlossenen Itemformaten von einer Missingquote zwischen 14 % und 21 %, bei offenen Items sogar von 31 % bis 53 %. Kaiser (2016) begründet dies mit der besonderen Schwierigkeit der Beantwortung von Vignettentests. So lässt sich der hohe Anteil an fehlenden Werten auch mit der kognitiven Belastung erklären. Es wird berichtet, dass bei Videovignetten von signifikant höherer kognitiver Belastung als bei Textvignetten auszugehen sei (Syring et al., 2015, S. 680). Kognitive Überforderung kann demnach auch als Unfähigkeit gewertet werden, adäquat mit belastenden Unterrichtssituationen umzugehen. Im Forschungssetting Seminar wurde dagegen eine signifikant höhere Angst beim Lösen von Textvignetten als beim Lösen von Videovignetten berichtet (Syring et al., 2015, S. 679).

Bei einem aus Video- und Textvignetten gemischten Test kann bei fehlenden Werten sowohl von kognitiver Überforderung als auch von Vermeidung von Angst ausgegangen werden. Da jedoch sowohl kognitive Fähigkeiten als auch die Überwindung von Angst vor dem Unterrichten Teil der Professionalität von Lehrkräften ist, könnte beides dazu legitimieren, auch bei Fällen mit vielen fehlenden Werten Missings mit 0 zu imputieren.

Um diese These zu überprüfen, wurde die Bedeutung der Imputation fehlender Werte mit 0 für den Datensatz in Biologie und Chemie systematisch untersucht. Hierfür wurden Stichprobenvergleiche auf Grundlage der theoretisch angenommenen Modelle des Noticing vorgenommen. Dieser Vergleich basiert auf der Hypothese, dass sich die Fitindices verbessern, wenn nur die besten Probandinnen und Probanden berücksichtigt werden und damit Fälle mit vielen imputierten Nullen entfernt werden. Bei der Berücksichtigung der an der Gesamtpunktzahl gemessenen besten 50 %, 40 %, 30 %, 20 % und 10 % der Probandinnen und Probanden werden die Fitindices sowohl beim Chemie- (vgl. Tabelle 5) als auch beim Biologietest (vgl. Tabelle 6) tendenziell schlechter.

Tabelle 5: Perzentilauswertung des Chemiedatensatzes

Fitindex	<i>N</i> = 50 %	<i>N</i> = 40 %	<i>N</i> = 30 %	<i>N</i> = 20 %	<i>N</i> = 10 %
RMSEA	0.053	keine Konvergenz	0.054	ERROR	ERROR
CFI/TLI	0.429/0.401	keine Konvergenz	0.448/0.421	ERROR	ERROR

Anmerkung: Nur die besten Probandinnen und Probanden wurden berücksichtigt. Mplus-basierte Berechnung mit Paarvergleichen (vgl. Kapitel 4.4).

Um auszuschließen, dass unter den 50 % mit besserer Gesamtpunktzahl nicht noch immer für das theoretische Modell zu inkompetente Probandinnen und Probanden enthalten waren, wurden diese in Zehnerschritten entfernt, sodass am Ende die besten 10 % übrig blieben (vgl. Tabelle 7 und Tabelle 8).

Tabelle 6: Perzentilauswertung des Biologiedatensatzes

Fitindex	$N = 50 \%$	$N = 40 \%$	$N = 30 \%$	$N = 20 \%$	$N = 10 \%$
RMSEA	0.044	0.043	0.039	ERROR	ERROR
CFI/TLI	0.542/0.530	0.532/0.520	0.482/0.469	ERROR	ERROR

Anmerkung: Nur die besten Probandinnen und Probanden wurden berücksichtigt. Mplus-basierte Berechnung mit Paarvergleichen (vgl. Kapitel 4.4).

Um sicherzustellen, dass die Verschlechterung der Fitindizes bei der alleinigen Berücksichtigung der besten Probandinnen und Probanden nicht allein an der evtl. zu geringen Stichprobe liegt, wurden künstlich die besten 10 %, 20 %, 30 % usw. der Probandinnen und Probanden zu den 50 % besten Probandinnen und Probanden addiert. Tabelle 7 ist zu entnehmen, dass auch dies nicht zu einer Verbesserung der Fitindizes des Chemietests führt, zumindest nicht bei CFI und TLI.

Tabelle 7: Mplus-basierte Berechnung mit Paarvergleichen (vgl. 4.4)

Fitindex	$N = 100 \%$ real	$N = 100 \%$ künstlich	$N = 90 \%$ künstlich	$N = 80 \%$ künstlich	$N = 70 \%$ künstlich	$N = 60 \%$ künstlich
RMSEA	0.070	0.061	Keine Kon- vergenz	0.059	0.058	Keine Kon- vergenz
CFI/TLI	0.816/ 0.807	0.361/ 0.330	Keine Kon- vergenz	0.396/ 0.367	0.416/ 0.388	Keine Kon- vergenz

Anmerkungen: Zu den 50 % der besten Probandinnen und Probanden des Chemiedatensatzes wurden in Zehnerschritten die besten Probandinnen und Probanden künstlich ergänzt, um auszuschließen, dass die Abhängigkeit von der Fähigkeit der Befragten nicht nur durch die zu geringe Stichprobe hervorgerufen ist. Vor dem Hintergrund der Ergebnisse im Vergleich zum realen Datensatz ($N = 100 \%$ real) kann eine solche Abhängigkeit zurückgewiesen werden.

Tabelle 8 zeigt beim Biologietest einen ähnlichen Effekt. Allerdings sind die Fitindices CFI und TLI nicht so weit voneinander entfernt ($N = 100 \%$ real vs. $N = 100 \%$ künstlich), wie dies beim Chemietest der Fall war, d. h. der künstliche und der reale Datensatz sind sich beim Biologietest ähnlicher als beim Chemietest.

Tabelle 8: Mplus-basierte Berechnung mit Paarvergleichen (vgl. 4.4)

Fitindex	$N = 100 \%$ real	$N = 100 \%$ künstlich	$N = 90 \%$ künstlich	$N = 80 \%$ künstlich	$N = 70 \%$ künstlich	$N = 60 \%$ künstlich
RMSEA	0.057	0.047	0.046	0.046	0.047	0.046

CFI/	0.669/	0.524/	0.532/	0.519/	0.549/	0.566/
TLI	0.660	0.512	0.520	0.506	0.537	0.554

Es wäre möglich, dass es sich bei den Befunden (Fitindices werden durch Imputation aller fehlenden Werte mit 0 besser) um ein Artefakt handelt. Hieraus ergibt sich die Hypothese:

„Durch die künstliche Vergrößerung²⁵ der Stichprobe werden die Fitindices stets verbessert“.

Um dies zu überprüfen, wird ein anderer Vignettentest-Datensatz (Naturwissenschaften Unterrichten Können (NUK): Brovelli et al., 2013) entsprechend dem Chemiedatensatz künstlich mit 7,9 % Fällen, die nur Items mit dem Punktwert 0 erhielten, ergänzt. Statt $N = 344$ enthielt der Datensatz $N = 371$ Fälle. Durch die künstliche Ergänzung der Fälle wurden die Fitindices deutlich schlechter²⁶ (RMSEA = 0.051, CFI/TLI = 0.795/0.614). Somit muss die Hypothese, dass es sich um ein Artefakt handelt, zurückgewiesen werden. Aus einem zurzeit nicht erklär-baren Grund passen sowohl beim Chemie- als auch beim Biologiedatensatz Fälle mit vielen imputierten Nullen am besten zur theoretisch angenommenen Struktur.

Es gibt weder einen einheitlichen noch einen „korrekten“ Umgang mit fehlenden Werten (vgl. Lüdtke et al., 2007). Aufgrund dieser Unklarheit über den korrekten Umgang mit fehlenden Werten wurden sowohl bei den Biologie- als auch den Chemie-Daten drei verschiedene Datensatzvarianten hinsichtlich des Umgangs mit fehlenden Werten berücksichtigt. Beim ersten Datensatz wurden im Sinne Rosts (1996, S. 381) alle Fälle, auch diejenigen, die keine Vignettenitems ausfüllten, berücksichtigt (im Weiteren „ $NA = 0$ “ genannt). Beim zweiten Datensatz wurden alle Fälle, die mehr als 50 % der Vignetten nicht ausfüllten, entfernt. Die restlichen fehlenden Werte wurden durch 0 ersetzt (im Weiteren „ $NA < 50 \% = 0$ “ genannt). Beim dritten Datensatz wurden wie schon beim zweiten Datensatz alle Fälle, die mehr als 50 % der Vignetten nicht ausfüllten, entfernt. Die verbliebenen fehlenden Werte wurden jedoch durch multiple Imputation mit SPSS 23 (IBM Corporation, 2015) unter Berücksichtigung aller Items, der Soziodemografie und der Kovariaten imputiert (im Weiteren „ $NA < 50 \% = imp$ “ genannt).

Da der MCAR-Test (Little, 1988) bei Berücksichtigung aller Missings in den Rohwerten (also auch Fällen, die nur fehlende Werte aufwiesen: $NA = 0$) beim Chemietest signifikant wurde ($p = .002$), auch wenn nur Fälle mit weniger als 50 % fehlende Werte ($NA < 50 \%$) berück-

²⁵ Mit „künstlicher Vergrößerung“ ist eine Vergrößerung der Stichprobe gemeint, die nicht auf realen Fällen fußt sondern die auf (teilweiser) Vielfältigkeit bestehender Fälle aus Stichproben beruht.

²⁶ Gerechnet mit dem „R-Package mirt“ (Chalmers (2012)).

sichtigt wurden ($p = .002$), sind die Chemiedatensätze $NA < 50 \% = imp$ eigentlich nicht aussagekräftig. Das heißt: Da die multiple Imputation direkt in die Rohwerte erfolgte, sind dort die Werte zumindest nur unter Vorbehalt zu interpretieren. Wenn der MCAR-Test jedoch direkt im bepunkteten Datensatz erfolgt, in dem Missings übertragen wurden, so gilt sowohl für den Datensatz $NA = 0$ ($p = .148$) als auch für den Datensatz $NA < 50 \%$ ($p = .157$) die MCAR-Annahme. Beim Biologiegesamttest gilt jedoch für die Rohwerte die MCAR-Annahme ($p = .349$), ebenso beim Datensatz $NA < 50 \%$ ($p = .371$). Ebenso gilt dies auch bei den bepunkteten Werten, bei denen Missings übertragen wurden ($NA = 0$: $p = .176$; $NA < 50 \%$: $p = .197$).

Da die Darstellung aller Ergebnisse der sechs Datensätze (drei Umgänge mit fehlenden Werten für den Biologie- und Chemietest) zu umfangreich und ggf. verwirrend ist, wird hauptsächlich der Datensatz $NA < 50 \% = 0$ berichtet. Bei der Diskussion der Ergebnisse wird jedoch auch auf die Ergebnisse der anderen Datensätze zurückgegriffen, um ein genaueres Bild vermitteln zu können, dies insbesondere im Sinne einer breiter abgestützten Kontrolle der Validität.

6 Ergebnisse

Im Folgenden werden zentrale Ergebnisse der entwickelten Tests berichtet. Zunächst wird der Test hinsichtlich der Reliabilität beschrieben. Dann werden einige Hinweise auf die Validität des Tests vorgestellt (Hypothese H2a-I bis H2b-II). Es folgt die Darstellung der zentralen Ergebnisse bezüglich Passung der theoretischen Struktur des Noticings und Reasonings jeweils in den Tests in Chemie und Biologie.

6.1 Testbeschreibung und Reliabilität

Die entstandenen Vignettentests über die Vermittlung von Modellkompetenz in den Schulfächern Chemie und Biologie wurden auf Grundlage des Paarvergleichs ausgewertet (vgl. Kapitel 4.4, S. 68). So entstandene Punktzahlen lassen sich wie in Tabelle 9 dargestellt zusammenfassen.

Tabelle 9: Übersicht über die durchschnittlichen Testergebnisse

Test	<i>N</i>	<i>M</i>	<i>SD</i>	Range	Theoretisches Maximum
Chemie	310	29.12	6.37	0-41	47
Biologie	348	52.61	9.93	10-75	84

Anmerkung: Als Grundlage wurde jeweils der Datensatz $NA < 50\% = 0$, bei dem Studierende mit weniger als 50 % fehlende Werte bei fehlenden Werten der Wert 0 zugewiesen wurde (vgl. Kapitel 5.3) genutzt.

Die Reliabilität gibt die Zuverlässigkeit einer Messung wieder. So kennzeichnet sie „den Grad der Genauigkeit, mit dem das geprüfte Merkmal gemessen wird“ (Bortz & Döring, 2006, S. 196). Das damit verfolgte Ziel ist ein möglichst kleiner Fehleranteil an der Messung.

Sowohl der Biologie- als auch der Chemiegesamttest (vgl. Tabelle 10) haben akzeptable bis gute Reliabilitätswerte (vgl. George & Mallery, 2003, S. 231).

Tabelle 10: Reliabilität der entwickelten Tests

Test	Cronbachs α (Itemanzahl)
Chemie	0,77 (69 Items)
Biologie _{gesamt}	0,83 (125 Items)
Biologie _{Vignetten}	0,81 (92 Items)
Biologie _{NoS-Modellvergleich}	0,63 (33 Items)

Anmerkung: KTT-basierte Berechnung von Cronbachs α mit SPSS 23 (IBM Corporation, 2015)

Allein der Cronbachs- α -Wert des Biologietests zu Modellunterscheidungen hinsichtlich NoS ist fragwürdig (George & Mallery, 2003, S. 231). Dies kann jedoch über die große Bandbreite

an Fragen zu Modellunterscheidungen und mit der Ähnlichkeit zum deklarativen Fachwissen erklärt werden. Auch in anderen Studien werden nicht oder nur knapp zufriedenstellende Reliabilitätswerte bei Fachwissenstests berichtet (Bernholt, Parchmann & Commons, 2009, S. 237: Verbrennungsprozesse: EAP/PV = .65; Kirschner, 2013, S. 104: geschlossene Items: $\alpha = .56$; Nehring, Nowak, Upmeyer zu Belzen & Tiemann, 2015, S. 1351: chemisches Konzeptwissen: $\alpha = .55$). Kirschner (2013) berechnet die Reliabilität schließlich nicht nur mit den geschlossenen, sondern zusätzlich mit offenen Items, um akzeptable α -Werte zu erreichen, wohingegen Nehring et al. (2015) zwar den nicht zufriedenstellenden Reliabilitätswert benennen, jedoch weiter mit dem Test rechnen und argumentieren. Insofern scheint eine Berücksichtigung des hier verwendeten Biologietests zu Modellunterscheidungen hinsichtlich NoS unter Vorbehalt gerechtfertigt.

6.2 Validitätshinweise

Die Gültigkeit einer Studie nachzuweisen, ist die schwierigste Aufgabe innerhalb der Bestimmung der Gütekriterien. Schließlich soll untersucht werden, „wie gut der Test in der Lage ist, genau das zu messen, was er zu messen vorgibt“ (Bortz & Döring, 2006, S. 200). Ein Nachweis ist unmöglich, bestenfalls können Hinweise vorgelegt werden (vgl. Weinert, 2001). Zu diesem Zweck wird Validität in drei Hauptaspekte unterteilt: Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität. Es gibt weitere Validitätsansätze (vgl. 4.3 Validierungsschritte), zu denen jedoch über statistische Verfahren nur schwer Hinweise gefunden werden können.

Die Inhaltsvalidität ist gegeben, wenn die Stützung der Annahme gelingt, dass die Testpunktzahl den Anspruch auf Domänenrelevanz und Repräsentativität erheben kann (Messick, 1989, S. 17) bzw. wenn Hinweise bestehen, dass genau das erhoben wird, was nach dem zugrundeliegenden Konstrukt inhaltlich logisch ist (Bortz & Döring, 2006, S. 200).

In ähnlicher Weise basiert die Kriteriumsvalidität auf dem Maß des empirischen Zusammenhangs zwischen dem Testscore und externen Faktoren wie Soziodemografie und Kovariaten, die dazu dienen, zu überprüfen, ob erwartbare Zusammenhänge in Form von Korrelationen bestehen (vgl. Bortz & Döring, 2006, S. 200; Messick, 1989, S. 17).

Zur Konstruktvalidität finden sich am schwersten Hinweise. Bei der Konstruktvalidität geht es um die Stützung der Behauptung, dass der Test das Konstrukt betrifft. Da Konstrukte jedoch oft offenen Konzepten ähneln und jede Operationalisierung einer Verarmung des Kon-

strukts entsprechen dürfte, ist eine Prüfung der Konstruktvalidität nur schwer und in Abwägung möglich (vgl. Messick, 1989, S. 17). Somit ist „[d]er Umstand, dass Testwerte so ausfallen, wie es die aus Theorie und Empirie abgeleiteten Hypothesen vorgeben“ (Bortz & Döring, 2006, S. 201) entscheidend.

6.2.1 Kovariaten als Hinweise auf Inhalts- und Kriteriumsvalidität

Um Hinweise auf die Kriteriumsvalidität des Tests zu erhalten, wurden neben der Soziodemografie auch Kovariaten erhoben (vgl. Kapitel 4.2, S. 62). Hintergrund dieses Vorgehens war der Validierungsaspekt *Aufgabengeneralisierbarkeit* (Messick, 1989). Theoretisch naheliegende, etablierte Tests zu diesem Bereich wurden hypothesengestützt miterhoben, um Hinweise auf die Eigenständigkeit des Vignettentests als auch des Zusammenhangs mit anderen angrenzenden Konstrukten zu erhalten. Die Reliabilitätswerte der Kovariaten (vgl. Kapitel 4.2, S. 62) in Abhängigkeit vom Datensatz (Chemie oder Biologie) sind Tabelle 11 zu entnehmen.

Tabelle 11: Reliabilitätswerte der Kovariaten

Autorinnen und Autoren	Test/Thema	N_{Items}	$N_{\text{Chemie/Biologie}}$	Cronbachs α: Chemie	Cronbachs α: Biologie
Vgl. Gerlitz und Schupp (2005)	BIG-Five/ Persönlichkeit	15	225/298	0,62	0,86
Retelsdorf et al. (2014)	ERBSE-L/ Berufsselbstkonzept	20	212/162	0,84	0,87
Pohlmann und Möller (2010)	FEMOLA/ Motivation Lehramtsstudium	33	217/286	0,87	0,84
Cacioppo und Petty (1982)	Need for Cognition/ Kognitionsbedürfnis	19	228/196	0,86	0,86
Kersting et al. (2008)	Wilde-Intelligenz-Test/Wortanalogien erkennen	20	127/162	0,77	0,81
König und Blömeke (2010)	Pädagogisches Unterrichtswissen/PPK-Test ²⁷	20	208/57	0,72	0,72
Witner und Tepner (2011c) und ETS (2015b)	Fachwissen Chemielehrkräfte/CK-Test	13	246/-	0,68	-
ETS (2015a)	Fachwissen Biologielehrkräfte/CK-Test	16	-/230	-	0,56

Anmerkung: KTT-basierte Berechnung von Cronbachs α mit SPSS 23 (IBM Corporation, 2015) in Abhängigkeit von den Gesamttests (Chemie oder Biologie).

²⁷ Nur geschlossene Items wurden genutzt, die offenen Items wurden entfernt.

Nach George und Mallery (2003) sind Reliabilitätswerte (Cronbachs α) größer 0,65 nutzbar. Dieser Wert ist weder beim Fachwissenstest Biologie, dessen Items aus ETS (2015a) entnommen und pilotiert wurden, noch beim Persönlichkeitstest der Chemiestichprobe (vgl. Gerlitz & Schupp, 2005) erreicht. Dennoch wird im Folgenden die Korrelation des Biologiescores mit dem CK-Test und die Korrelation des Chemiescores mit dem Persönlichkeitstest (vgl. Gerlitz & Schupp, 2005) berichtet. Grund dafür ist, dass zum einen Wissenstests häufiger schlechtere Reliabilitätswerte aufweisen (Bernholt et al., 2009, S. 237: Verbrennungsprozesse: EAP/PV = .65; Kirschner, 2013, S. 104: geschlossene Items: α = .56; Nehring et al., 2015, S. 1351: chemisches Konzeptwissen: α = .55) und es sich zum anderen beim häufig erprobten Persönlichkeitstest BIG-FIVE, abweichend von anderen Studienergebnissen, um einen Ausreißer zu handeln scheint. Diese Ergebnisse dürfen jedoch nur mit besonderer Vorsicht interpretiert werden. Weiterhin fällt auf, dass trotz der unterschiedlichen Stichproben aus unterschiedlichen Grundgesamtheiten in Chemie und Biologie häufig ähnliche Reliabilitätswerte gemessen wurden. Alle anderen Tests können bezüglich ihrer Reliabilität als akzeptabel bis gut bezeichnet werden (George & Mallery, 2003, S. 231), der CK-Test Chemie als annähernd akzeptabel. Allerdings sind diese Reliabilitätswerte der CK-Tests stabil: Zieht man die Stichprobe des EKoL-Gesamtprojekts (vgl. Fußnote 3, S. 14) zurate, so erhält der Biologie-CK-Test ein Cronbachs α von .55 ($N = 544$) statt wie in der hier vorgestellten Studie .56 ($N = 246$). Der Cronbachs- α -Wert ist im EKoL-Gesamtprojekt ($\alpha = .68$ ($N = 464$)) identisch mit dem Wert der hier vorgestellten Studie ($\alpha = .68$ ($N = 230$)). Zumindest beim Biologietest wäre also eine weitere Erhebung mit einem reliablen CK-Test wünschenswert, um die Hinweise auf die Korrelation zum Vignettenscore erhärten zu können.

Im Kapitel 4.2 (S. 62) wurde der Grund für den Einsatz der Kovariaten sowie die erwarteten Ergebnisse kurz umrissen. Nun folgt die Darstellung der manifesten Korrelationen (gerechnet mit SPSS 23: IBM Corporation, 2015). In Tabelle 12 sind die manifesten Korrelationen zum Chemietestscore zu entnehmen.

Tabelle 12: Manifeste Korrelation zwischen Chemietestscore und Kovariaten bzw. deren Subskalen

Kovariate		Gesamtscore
Persönlichkeit: Offenheit	<i>r</i>	-.16*
	<i>p</i>	.016
	<i>N</i>	235
Motivation: Geringe Schwierigkeit des Lehramtsstudiums	<i>r</i>	-0.24***
	<i>p</i>	<.001

	<i>N</i>	214
Motivation: Soziale Einflüsse	<i>r</i>	-.22***
	<i>p</i>	.001
	<i>N</i>	214
Intelligenz/Wortanalogien	<i>r</i>	.33***
	<i>p</i>	<.001
	<i>N</i>	259
PK-Test	<i>r</i>	.39***
	<i>p</i>	<.001
	<i>N</i>	241
CK-Test Chemie	<i>r</i>	.32***
	<i>p</i>	<.001
	<i>N</i>	292

Anmerkungen: Aufgrund fehlender Werte bei den Kovariaten unterscheiden sich die Stichprobengrößen (*N*).
r: Pearsons Korrelationskoeffizient; * $p \leq .05$; *** $p \leq .001$.

Beim Biologietest wird nicht nur der Vignettentestscore korreliert, sondern auch zusätzlich der NoS-Wissenstest zur Modellunterscheidung. Daher wird hier ein Vignettenscore und ein NoS-Score dem Gesamtscore, der Vignetten- und NoS-Score zusammenfasst, gegenübergestellt. Alle signifikanten Korrelationen zu diesen drei Punktzahlen können Tabelle 13 entnommen werden.

Tabelle 13: Manifeste Korrelation zwischen den Biologiescores und Kovariaten bzw. deren Subskalen

Kovariate/Score		Gesamtscore	Vignetten- score	NoS-Score
Gesamtscore	<i>r</i>	1	.92***	.72***
	<i>p</i>		<.001	<.001
	<i>N</i>	348	348	348
Vignettenscore	<i>r</i>	.92***	1	.40***
	<i>p</i>	<.001		<.001
	<i>N</i>	348	348	348
NoS-Score	<i>r</i>	.72***	.40***	1
	<i>p</i>	<.001	<.001	
	<i>N</i>	348	348	348
Persönlichkeit: Verträglichkeit	<i>r</i>	0.11	.15*	-0.01
	<i>p</i>	.066	.011	.838
	<i>N</i>	306	306	306
Motivation: Fachliches Interesse	<i>r</i>	.12*	.13*	0.04
	<i>p</i>	.040	.022	.443
	<i>N</i>	300	300	300
Motivation: Geringe Schwierigkeit des Lehramtsstudiums	<i>r</i>	-.20***	-.22***	-0.09
	<i>p</i>	<.001	<.001	.135
	<i>N</i>	301	301	301
Motivation: Pädagogisches Interesse	<i>r</i>	.13*	.16**	0.02
	<i>p</i>	.029	.006	.769
	<i>N</i>	300	300	300
Motivation: Soziale Einflüsse	<i>r</i>	-.14*	-.15**	-0.06
	<i>p</i>	.017	.008	.340
	<i>N</i>	301	301	301
Need for Cognition	<i>r</i>	.18**	.19**	0.10
	<i>p</i>	.009	.007	.152

	<i>N</i>	201	201	201
Intelligenz/ Wortanalogien	<i>r</i>	.37***	.36***	.23***
	<i>p</i>	<.001	<.001	<.001
	<i>N</i>	324	324	324
PK-Test	<i>r</i>	.25***	.22***	.19**
	<i>p</i>	<.001	.001	.004
	<i>N</i>	218	218	218
CK-Test Biologie	<i>r</i>	.41***	.36***	.34***
	<i>p</i>	<.001	<.001	<.001
	<i>N</i>	323	323	323

Anmerkungen: Aufgrund fehlender Werte bei den Kovariaten unterscheiden sich die Stichprobengrößen (*N*).
r: Pearsons Korrelationskoeffizient; * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

Im Folgenden werden die Korrelationen zwischen den Testscores des Chemie- und Biologietests sowie den Kovariaten diskutiert. Wie erwartet findet sich zu den Fachwissenstests und den Scores in Chemie und Biologie mit die stärkste, jedoch moderat-signifikante Korrelation (Chemie: $r = .32$; Biologie_{Gesamtscore}: $r = .41$; Biologie_{Vignettentest}: $r = .36$; Biologie_{NoS-Modellunterscheidungen}: $r = .34$). Ebenfalls erwartungskonform kann die Bedeutung des Fachwissens für die Fachdidaktik (hier: Vermittlung von Modellkompetenz) repliziert werden (Blömeke, Seeber et al., 2008, S. 69, $r = .81$; Voss et al., 2011, S. 963, latente Korrelation = $.91^*$). Allerdings kann bei unserem Vignettentest erwartungskonform PCK hinsichtlich *Vermittlung von Modellkompetenz* als eigenständiges Konstrukt bezeichnet werden, während bei den bisherigen Largescale-Studien (Blömeke, Seeber et al., 2008; Voss & Kunter, 2011) auch aufgrund des Testformats von einer immens großen Nähe ausgegangen wurde und PCK und CK als kaum trennbar galten.

Ebenfalls zeigt sich eine im Vergleich zu den anderen Korrelationen hohe aber moderate Korrelation zwischen den Testscores Biologie und Chemie und der Intelligenz (Chemie: $r = .33$; Biologie_{Gesamtscore}: $r = .37$; Biologie_{Vignettentest}: $r = .36$; Biologie_{NoS-Modellunterscheidungen}: $r = .23$). Auch dies bestätigt die Annahme, dass zum adäquaten Umgang mit Modellen abstraktes Denken, welches Intelligenz und sprachliche Fähigkeiten zum Verständnis der Textvignetten voraussetzt, nötig ist. Die schwach- bis moderat-signifikante Korrelation (Chemie: $r = .39$; Biologie_{Gesamtscore}: $r = .25$; Biologie_{Vignettentest}: $r = .22$; Biologie_{NoS-Modellunterscheidungen}: $r = .19$) zum PK-Test (König & Blömeke, 2010) war ebenfalls erwartet worden und findet sich beispielsweise bei Voss et al. (2011, S. 963: latente Korrelation = $.42^*$) ausgeprägter.

Die FEMOLA-Skala (Pohlmann & Möller, 2010) *geringe Schwierigkeit des Lehramtsstudiums* korreliert schwach negativ mit dem Chemie_{vignettengesamtscore} ($r = -.24$) und dem Biologie_{gesamtscore} ($r = -.20$) sowie dem Biologie_{vignettenscore} ($r = -.22$), nicht jedoch zum NoS-Score. Dies war nicht erwartet worden. Allerdings zeigt das negative Vorzeichen, dass

eher Probandinnen und Probanden, die nicht wegen geringer Schwierigkeit des Lehramtsstudiums Lehrkraft werden wollen, gut abschneiden.

Ausschließlich beim Chemietest findet sich eine schwach negative Korrelation zum Persönlichkeitsmerkmal Offenheit (vgl. Gerlitz & Schupp, 2005; $r = -.16$). Dies war so nicht erwartet worden. Lehramtsstudierende mit guten Vignettentestleistungen scheinen tendenziell ein wenig verschlossener zu sein.

Korrelationen ausschließlich zwischen Biologiescores und Kovariaten (und nicht zum Chemiescore) finden sich unerwarteterweise zwischen dem Vignettenscore und der Skala Persönlichkeitseigenschaft „Verträglichkeit“ als Teil des BIG-5-Fragebogens (vgl. Gerlitz & Schupp, 2005; $r = .15$). Eine schwach-positive Korrelation zwischen fachlichem Interesse (Biologie_{Gesamtscore}: $r = .12$; Biologie_{Vignettentest}: $r = .13$) und pädagogischem Interesse (Biologie_{Gesamtscore}: $r = .13$; Biologie_{Vignettentest}: $r = .16$) (Pohlmann & Möller, 2010) war zwar nicht erwartet worden, ist jedoch interpretierbar. Die schwache Korrelation zwischen Gesamt- bzw. Vignettenscore und dem Kognitionsbedürfnis (Cacioppo & Petty, 1982) ist erwartungskonform (Biologie_{Gesamtscore}: $r = .18$; Biologie_{Vignettentest}: $r = .19$).

Der Fall, dass erwartete Korrelationen zwischen den Scores und den Kovariaten nicht gemessen wurden, trat nicht ein. Die Korrelationsergebnisse beziehen sich auf den Datensatz $NA < 50 \% = 0$. Die Ergebnisse mit den Datensätzen $NA = 0$ und $NA < 50 \% = imp$ unterscheiden sich jedoch nur geringfügig. Insgesamt werden die Ergebnisse als Hinweis auf die Inhalts- und Kriteriumsvalidität interpretiert.

6.2.2 Soziodemografische Hinweise auf Inhalts- und Kriteriumsvalidität

Soziodemografische Angaben können wichtige Hinweise auf die Validität eines Tests liefern. Erwartet wird kein bzw. ein schwacher Zusammenhang zum Alter. Das Alter sollte nur gering korrelieren, um der Kriteriumsvalidität gerecht zu werden. Sollte die Kompetenz eine starke Abhängigkeit zum Alter zeigen, würde das Alter etwas über die Fähigkeit der Vermittlung von Modellkompetenz an Schülerinnen und Schüler aussagen – das wäre nicht erwartungskonform. Allerdings könnte das Alter über Lehrerfahrung moderiert sein, sodass Zusammenhänge möglich sind. Eine schwache bis moderate negative²⁸ Korrelation mit der Durchschnittsnote aus dem Abitur wird erwartet, da Abitur und Intelligenz zusammenhängen (vgl.

²⁸ Wegen des deutschen Notensystems, bei dem 1 die beste und 4 die schlechteste Abiturnote ist, die als bestanden gilt, wird eine negative Korrelation erwartet.

z. B. Rindermann & Neubauer, 2000) und der abstraktes Denken fordernde Umgang mit Modellen hiervon wahrscheinlich beeinflusst wird. Tabelle 14 sind die Korrelationen des Chemiescores und der verschiedenen Biologiescores mit dem Alter und der Abiturnote der Probandinnen und Probanden zu entnehmen.

Tabelle 14: Manifeste Korrelationen zwischen Alter und Abiturnote und den Chemie- und Biologiescores

		Chemiegesamt- score	Biolo- giegesamt- score	Biologievignet- tenscore	Biologie-NoS- Score
Alter	<i>r</i>	0.06	0.14	0.08	.19
	<i>p</i>	.353	.057	.281	.009
	<i>N</i>	266	183	183	183
Abiturnote	<i>r</i>	-0.04	-.23***	-.22***	-.15*
	<i>p</i>	.484	.001	.001	.027
	<i>N</i>	266	209	209	209

Anmerkungen: Aufgrund fehlender Werte bei den soziodemografischen Angaben unterscheiden sich die Stichprobengrößen (*N*).

r: Pearsons Korrelationskoeffizient; * $p \leq .05$.; *** $p \leq .001$

Entsprechend der Annahmen korreliert Alter nicht mit den Scores bzw. nur schwach mit dem Wissen um Modellunterscheidungen (NoS). Erwartungskonform ergeben sich schwache negative Korrelationen zwischen den Biologiescores und der Abiturnote (d. h., dass Biologielehramtsstudierende mit besserem Abitursdurchschnitt tendenziell bessere Scores haben, da die 1 die beste und 4 die schlechteste im Abiturszeugnis möglich Note ist). Der Chemiescore korreliert entgegen der Erwartung nicht mit der Abiturnote.

Die Testscores sollten unabhängig vom Geschlecht sein, da Vermittlung von Modellkompetenz geschlechterunabhängig sein sollte. Die Ergebnisse bestätigen diese Einschätzung: Das Geschlecht hat weder beim Chemiegesamt-score ($p = .545$, $F(1, 296) = 0.37$) noch beim Biologiesamt-score ($p = .28$, $F(1, 330) = 1.20$) einen signifikanten Einfluss.

Zwischen den Befragten der verschiedenen Studiengänge sollten Unterschiede zu finden sein. Häufig haben bei PCK-Tests Lehramtsstudierende des Gymnasiums bzw. Lehrkräfte aus dem Gymnasium signifikant die durchschnittlich höchsten Punktzahlen (Kunter & Voss, 2011, S. 99; Riese & Reinhold, 2010). Studierende, die kein Lehramtsstudium studieren, sollten schlechter abschneiden als Lehramtsstudierende (vgl. z. B. Brovelli et al., 2013, 321 f.).

Beim Chemietest zeigt sich post hoc jedoch kein signifikanter Unterschied. Marginal signifikant war nur der Unterschied gymnasiales Lehramtsstudium vs. kein Lehramtsstudium ($p = .064$, $F(5, 237) = 1.88$). Möglicherweise hängt das bessere Abschneiden der Lehramtsstudierenden für Gymnasien mit der stärkeren fachwissenschaftlichen Ausrichtung des Studiums zusammen. Dass Fachwissen allein nicht ausreicht für die erfolgreiche Beantwortung des

Vignettentests zeigt sich darin, dass die Chemiestudierenden ohne Lehramtsstudium zwar marginal, aber signifikant schlechter abschnitten, obwohl ein höheres Fachwissen erwartbar wäre. Beim Biologietest finden sich entgegen der Erwartung post hoc keine signifikanten Unterschiede zwischen den Studiengängen.

Entscheidend für die Bewertung des Tests hinsichtlich der Validität ist die Entwicklung der Testscores in Abhängigkeit von den geleisteten Unterrichtspraktikumsstunden und den Fachsemestern. Bezüglich der Unterrichtspraktikumsstunden wird ein kontinuierlicher signifikanter Anstieg erwartet, sowohl für die Tests zur Vermittlung von Modellkompetenz im Chemie- als auch im Biologieunterricht. Abbildung 14 kann die Bestätigung dieser Annahme für den Chemietest ($p = .020$, $F(2, 246) = 3.94$) entnommen werden. Hypothese H2b-II („Der Testscore im Chemietest steigt querschnittlich in Abhängigkeit von der Anzahl der geleisteten Unterrichtsstunden an.“) muss damit *nicht falsifiziert* werden.

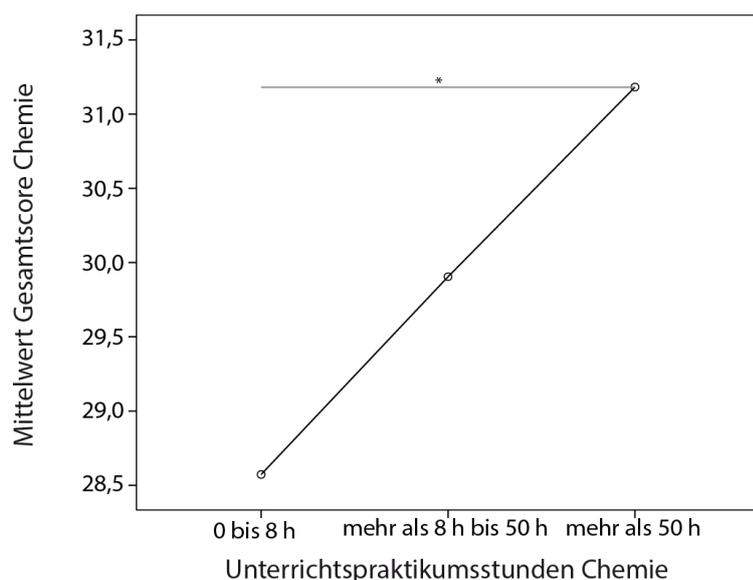


Abbildung 14: Querschnittliche Entwicklung des Chemietestscores über die gruppierten Unterrichtspraktikumsstunden. Es zeigen sich signifikante ($p = .020$, $F(2, 246) = 3.94$) Unterschiede zwischen bis zu acht Stunden und mehr als 50 Stunden Unterrichtserfahrung, eigene Darstellung.

Erwartungswidrig gilt dies nicht für den Biologiegesamttestscore (Vignetten- und NoS-Testscore). Hier ist kein signifikanter Unterschied bzw. Anstieg hinsichtlich Unterrichtspraktikumsstunden feststellbar. Hypothese H2b-I („Der Testscore im Biologietest steigt querschnittlich in Abhängigkeit von der Anzahl der geleisteten Unterrichtsstunden an.“) muss also zurückgewiesen bzw. *falsifiziert* werden.

Bezüglich der Fachsemester wird erwartet, dass sowohl im Chemie- als auch im Biologietest ein kontinuierlicher und signifikanter Anstieg bis zum letzten Semester und ebenfalls ein Anstieg des Testscores bei Lehramtsanwärterinnen und Lehramtsanwärttern, hier Referendariat genannt, gemessen werden kann. In Abbildung 15 ist die Entwicklung des durchschnittlichen Chemietestscores in Abhängigkeit vom Fachsemester dargestellt.

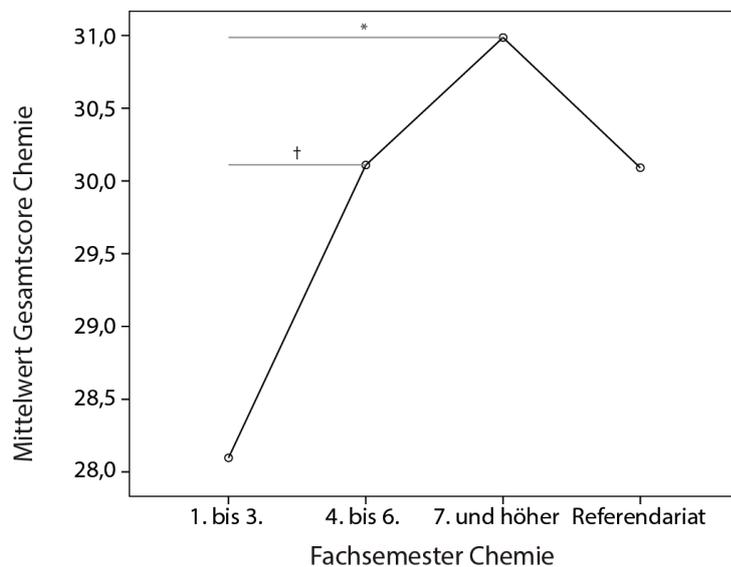


Abbildung 15: Querschnittliche Entwicklung des Chemietestscores über gruppierte Fachsemester. Es zeigen sich marginal signifikante (\dagger : $p = .095$, $F(3, 268) = 3.46$) Unterschiede zwischen 1.-3. Fachsemester und 4.-6. Fachsemester sowie signifikante ($p = .013$, $F(3, 268) = 3.46$) Unterschiede zwischen 1.-3. Fachsemester und 7.-8. Fachsemester, eigene Darstellung.

Im Chemietest zeigen sich ein marginal signifikanter sowie ein signifikanter durchschnittlicher Testscoreanstieg zwischen Studiumsanfängerinnen und Studiumsanfängern und Studierenden aus dem vierten bis sechsten sowie dem siebten bis achten Fachsemester. Es folgt ein nichtsignifikanter Abfall vom Ende des Studiums hin zum Referendariat. Eine mögliche Begründung für diesen nicht erwarteten Abfall hin zur Zeit des Referendariats ist wahrscheinlich in der kleinen Substichprobe zu suchen: Nur $n = 11$ Personen gaben an, dass sie sich im Referendariat befänden. Insofern hat dieser nichtsignifikante Abfall keine Vorhersagekraft für eine Kompetenzabnahme im Referendariat im Vergleich zum Ende des Lehramtsstudiums. Hypothese H2a-II („Der Testscore aus dem Chemietest steigt querschnittlich in Abhängigkeit von der Anzahl der Fachsemester an.“) muss somit *nicht falsifiziert* werden. Über die Entwicklung zwischen Studium und Referendariat kann wegen der geringen Angaben keine Aussage gemacht werden.

Im Biologietest zeigt sich auch im nichtgruppierten Querschnittsverlauf des Gesamttestscores ein signifikanter Anstieg zwischen Erstsemestern und Probandinnen und Probanden im Refe-

rendariat, obwohl die Substichprobe ($n = 18$) der Lehramtsanwärterinnen und Lehramtsanwärter im Vergleich zum Chemietest nur unwesentlich größer ist (vgl. Abbildung 16).

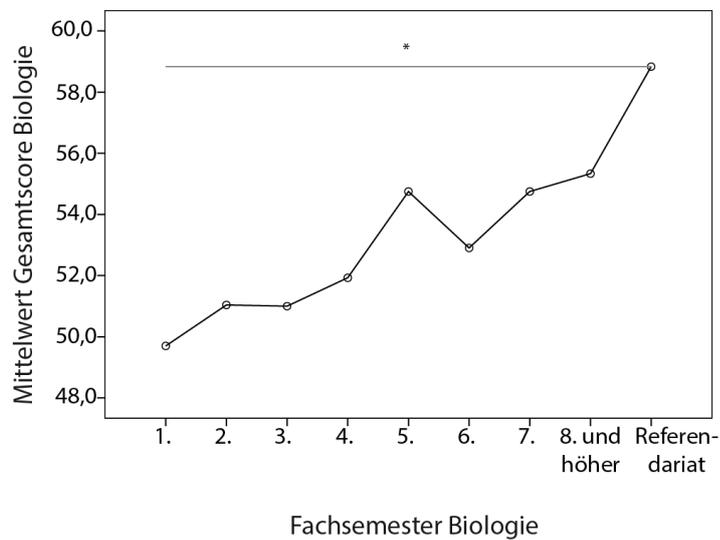


Abbildung 16: Querschnittliche Entwicklung des Biologiegesamttestescores über die Fachsemester. Es zeigen sich signifikante ($p = .034$, $F(8, 205) = 1.94$) Unterschiede zwischen erstem Fachsemester und Referendariat, eigene Darstellung.

Gruppiert man wie schon beim Chemietest auch beim Biologietest die Fachsemester, so kommt es erwartungskonform zu dem in Abbildung 17 dargestellten kontinuierlichen und signifikanten Anstieg des durchschnittlichen Testscores zwischen Studierenden am Anfang des Studiums und am Ende des Studiums sowie zu einem hochsignifikanten Anstieg hin zum Referendariat.

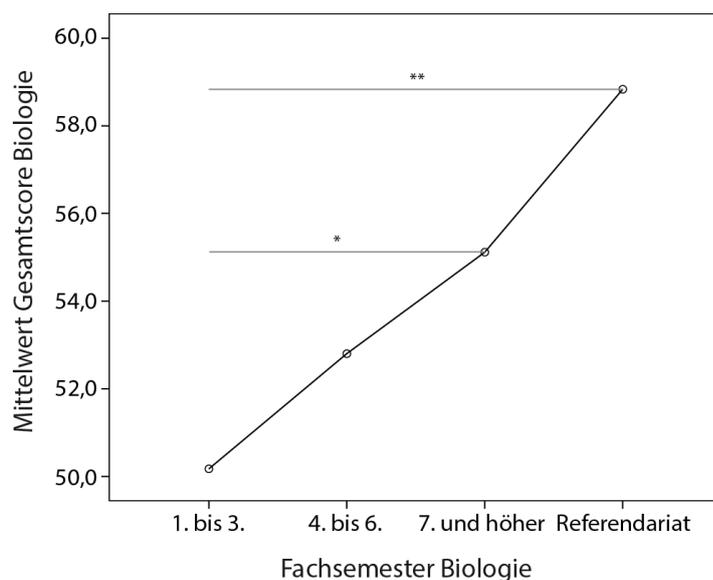


Abbildung 17: Querschnittliche Entwicklung des Biologiegesamttestescores über die gruppierten Fachsemester. Es zeigen sich signifikante ($p = .048$, $F(3, 210) = 5.05$) Unterschiede zwischen dem ersten bis dritten Fach-

mester und dem siebten bis achten Semester und hochsignifikante ($p = .006$, $F(3, 210) = 5.05$) Unterschiede zwischen dem ersten bis dritten Fachsemester und dem Referendariat, eigene Darstellung.

Hypothese H2a-I („Der Testscore aus dem Biologietest steigt querschnittlich in Abhängigkeit von der Anzahl der Fachsemester an.“) muss wie erwartet nicht falsifiziert werden. Im Gegensatz zum Chemietest gilt dies sogar bis hin zur Zeit des Referendariats.

Zusammenfassend liefert die Entwicklung über die Anzahl der geleisteten Unterrichtsstunden sowie über die Fachsemester wichtige Indizien für die Validität des Chemietests. Aufgrund der geringen Stichprobe kann kein signifikanter Anstieg hin zum Referendariat berichtet werden. Diesen Validitätshinweis kann zwar der Biologietest liefern, nicht jedoch der Anstieg des Gesamtestscores über die geleisteten Unterrichtsstunden.

6.2.3 Test- und Itemschwierigkeiten

Bei der Entwicklung eines Tests ist es interessant, wie schwierig dieser insgesamt und im Speziellen die einzelnen Items sind. Nur so kann Hypothese H3-II („Die Testschwierigkeit liegt unabhängig vom Umgang mit fehlenden Werten im mittleren Schwierigkeitsbereich.“) falsifiziert oder gegebenenfalls nicht falsifiziert werden. Abbildung 18 zeigt exemplarisch die Häufigkeit, mit der die Paarvergleiche im Biologietest entsprechend dem Expertinnen- und Expertenurteil gelöst wurden.

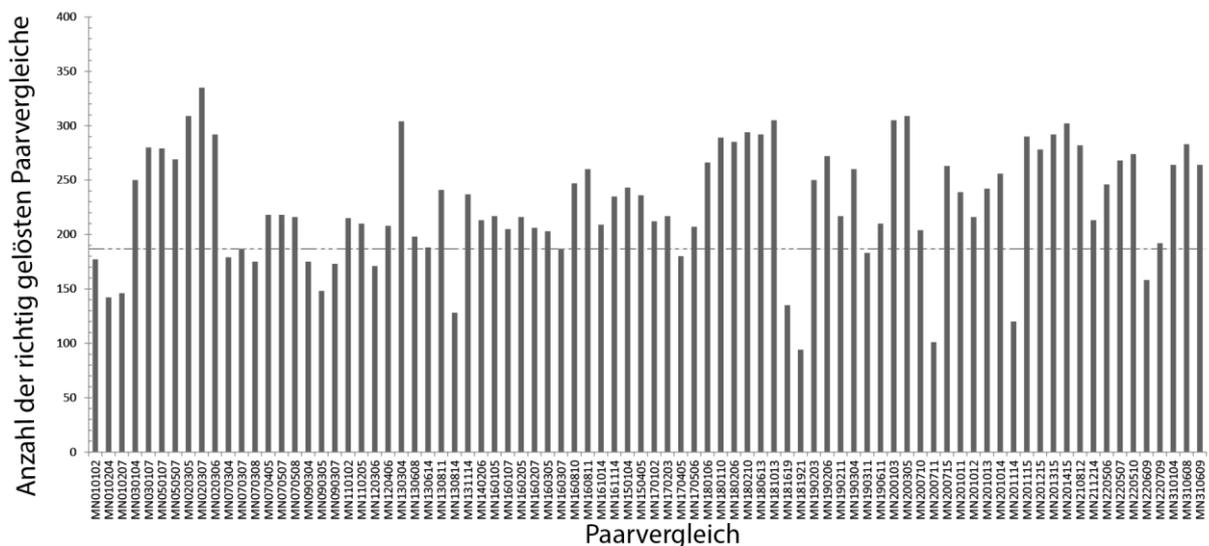


Abbildung 18: Biologie-Vignettest und -NoS-Test (Datensatz $NA = 0$): Anzahl der richtig gelösten (Pseudo-) Items bzw. Paarvergleiche. Die gestrichelte Linie entspricht 50 % der Probandinnen und Probanden ($N = 188$), eigene Darstellung.

Eine detaillierte Angabe der Schwierigkeit des gesamten Biologietests ist über Abbildung 18 nur bedingt möglich. Aussagen über die Schwierigkeit eines Items in Abhängigkeit von der Fähigkeit einer Person sind in dieser Form auf Grundlage der klassischen Testtheorie (KTT) nicht möglich. Für Fragestellungen wie diese wird häufig die Item-Response-Theorie (IRT) verwendet. Diese probabilistische Testtheorie erlaubt Aussagen über die Wahrscheinlichkeit, mit der ein Proband oder eine Probandin mit einer spezifischen Personenfähigkeit (θ) ein Item einer bestimmten Schwierigkeit lösen kann (vgl. Edelen & Reeve, 2007). Hieraus ergibt sich bei jedem Test ein sogenannter Personenfähigkeitsbereich. Innerhalb dieses Bereichs lassen sich aufgrund der erhöhten Häufigkeiten der Items genauere Aussagen treffen als außerhalb. Der Test ist dort also reliabler als an den Rändern.

Zur Veranschaulichung der Testschwierigkeit in Abhängigkeit von der Itemschwierigkeit wird in Abbildung 19 bis Abbildung 24 jeweils die IRT-basierte Total Information Curve (TIC) präsentiert. Sie wurde auf der Grundlage des R-Packages *mirt* (Chalmers, 2014) berechnet und dann um die Itemschwierigkeiten ergänzt.

Bei Berücksichtigung aller Probandinnen und Probanden ergab sich Abbildung 19:

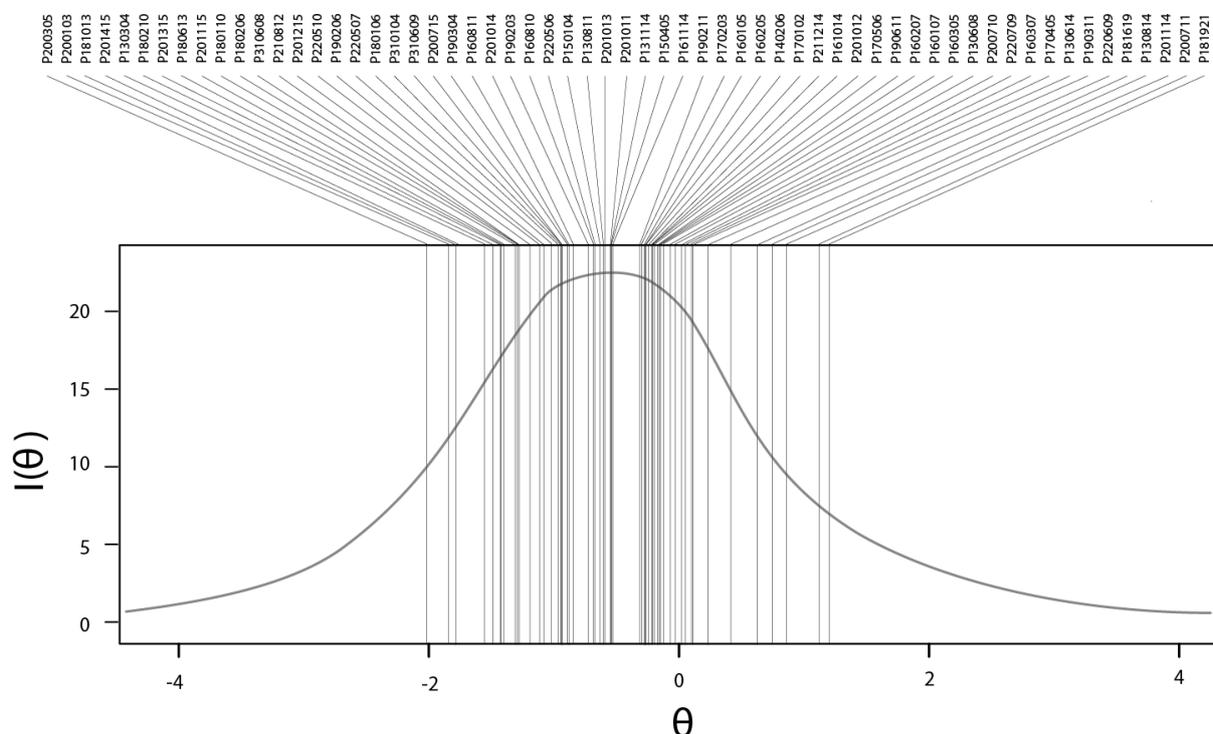


Abbildung 19: Testschwierigkeit des Biologievignetentests in Abhängigkeit von der Itemschwierigkeit (Datensatz: $NA = 0$, $N = 376$, ohne Modellvergleichs-/NoS-Items), berechnet mit dem R-Package *mirt* (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. P steht für den Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung.

Die mit dem R-Package *mirt* (Chalmers, 2014) berechnete Total Information Curve wurde mit dem Grafikprogramm Adobe Illustrator CS6 (Adobe, 2012) um die senkrecht abgetragenen Items in Abhängigkeit von ihrer jeweiligen Schwierigkeit ergänzt. Bei den senkrecht abgetragenen sogenannten Pseudoitems handelt es sich um die Paarvergleiche (P). Deren Schwierigkeit ergibt sich aus der Spiegelung der Itemleichtigkeit (d), also dem Drehen der Vorzeichen von minus in plus und umgekehrt. Die ersten beiden Zahlen kodieren die Vignettennummer, die letzten vier Ziffern den Paarvergleich. So ist P200305 (Vignette 20 mit dem Paarvergleich zwischen den Items 03 und 05) besonders einfach, P181921 hingegen besonders schwer.

Der Biologievignettentest ist bei Verwendung des Datensatzes $NA = 0$ nur etwas leichter (Wendepunkt der Kurve) als die durchschnittliche Fähigkeit der befragten Personen ($\theta = 0$).

Abbildung 20 zeigt die TIC des Datensatzes $NA < 50 \% = 0$. Die Personenfähigkeit (θ) in Abhängigkeit von der Testschwierigkeit unterscheidet sich kaum vom Datensatz $NA = 0$, allein der Informationsgehalt ($I(\theta)$) ist geringer.

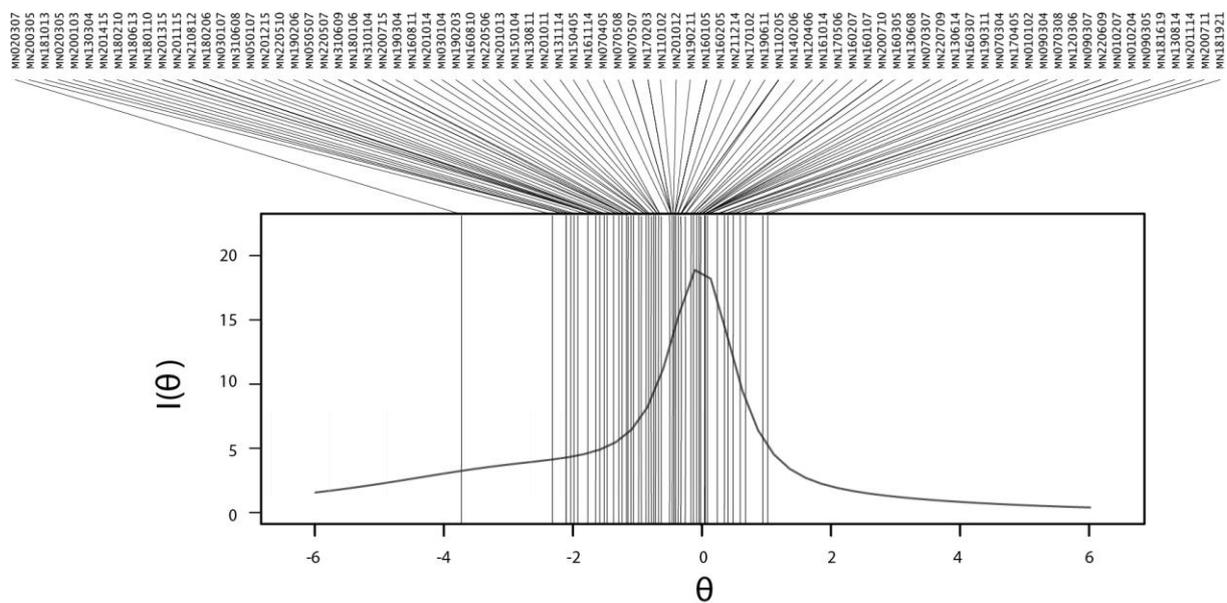


Abbildung 20: Testschwierigkeit des Biologievignettentests in Abhängigkeit von der Itemschwierigkeit (Datensatz: $NA < 50 \% = 0$, $N = 348$), berechnet mit dem R-Package *mirt* (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. MN steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung.

Der Datensatz $NA < 50 \% = imp$ (vgl. Abbildung 22) erzeugt ebenfalls eine ähnliche Testschwierigkeitskurve wie schon die Datensätze $NA = 0$ und $NA < 50 \% = 0$. Allerdings ist der Test in diesem Fall geringfügig schwieriger und der Informationsgehalt ($I(\theta)$) deutlich geringer als bei den beiden anderen Datensätzen.

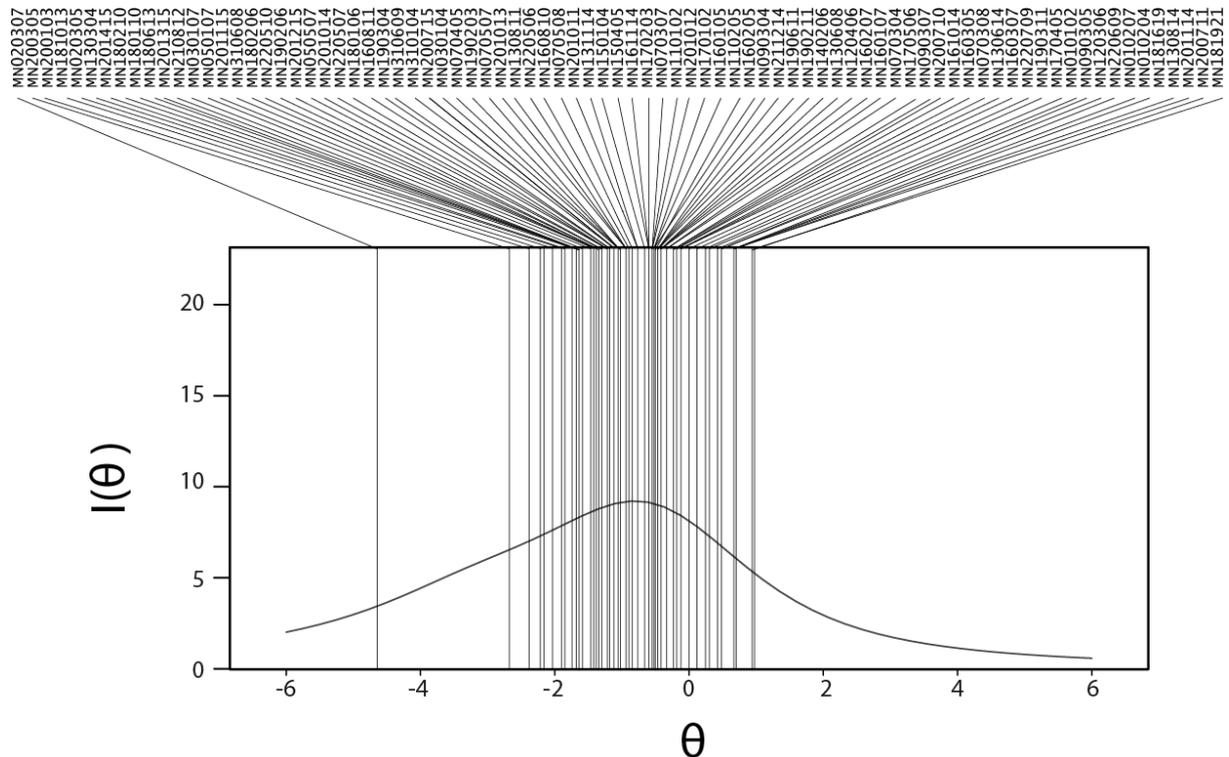


Abbildung 21: Testschwierigkeit des Biologievignettentests in Abhängigkeit von der Itemschwierigkeit (Datensatz: $NA < 50\% = imp$, $N = 348$), berechnet mit dem R-Package *mirt* (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. MN steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung.

Zusammenfassend kann für den Biologietest Folgendes festgestellt werden: Es gibt zwar Unterschiede in der Testschwierigkeit, aber die Abweichungen sind relativ gering. Imputiert man die fehlenden Werte von Personen, die mindestens 50 % des Tests ausgefüllt haben, wird der Test etwas leichter als wenn man diese auf 0 setzt bzw. als wenn man alle fehlenden Werte durch 0 ersetzt. Der Informationsgehalt nimmt umgekehrt, also vom Datensatz $NA = 0$ hin zu $NA < 50\% = imp$, ab.

Auf Grundlage dieser Ergebnisse ist von einer recht guten Passung der Schwierigkeit für Lehramtsstudierende des Faches Biologie auszugehen, da der Test eher einfach als schwierig ist (Wendepunkt der Kurve bei $\theta < 0$). Hypothese H3-II („Die Testschwierigkeit liegt unabhängig vom Umgang mit fehlenden Werten im mittleren Schwierigkeitsbereich.“) muss wegen nur geringer Unterscheidung der Testschwierigkeit in Abhängigkeit vom Datensatz bzw. Umgang mit fehlenden Werten *nicht falsifiziert* werden. Allerdings liegt hierzu auch kein Schwellenwert vor.

Für die Testschwierigkeitsanalyse des Chemietests wurde die gleiche Vorgehensweise gewählt. Abbildung 22 zeigt beim Datensatz $NA = 0$ eine mit dem Biologietest vergleichbare geringe Testschwierigkeit mit hohem Informationsgehalt ($I(\theta)$).

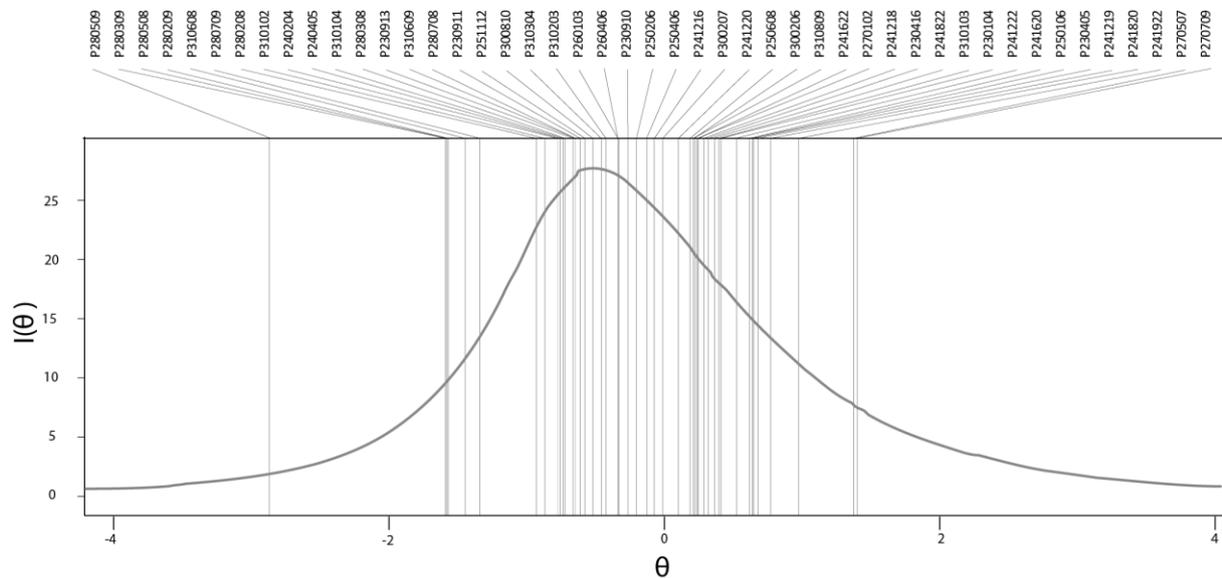


Abbildung 22: Testschwierigkeit des Chemievignetentests (Datensatz $NA = 0$) in Abhängigkeit von der Itemschwierigkeit, berechnet mit dem R-Package `mirt` (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. P steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung.

Ähnliches gilt für Abbildung 23. Allerdings ist der Informationsgehalt hier auch wieder deutlich geringer als bei $NA = 0$, der Test wird etwas schwieriger. Der Abbildung 24 ist zu entnehmen, dass zwischen den Chemiedatensätzen $NA < 50\% = 0$ und $NA < 50\% = imp$ kaum Unterschiede existieren. Dies gilt sowohl bezüglich der Schwierigkeit als auch bezüglich des geringen Informationsgehaltes.

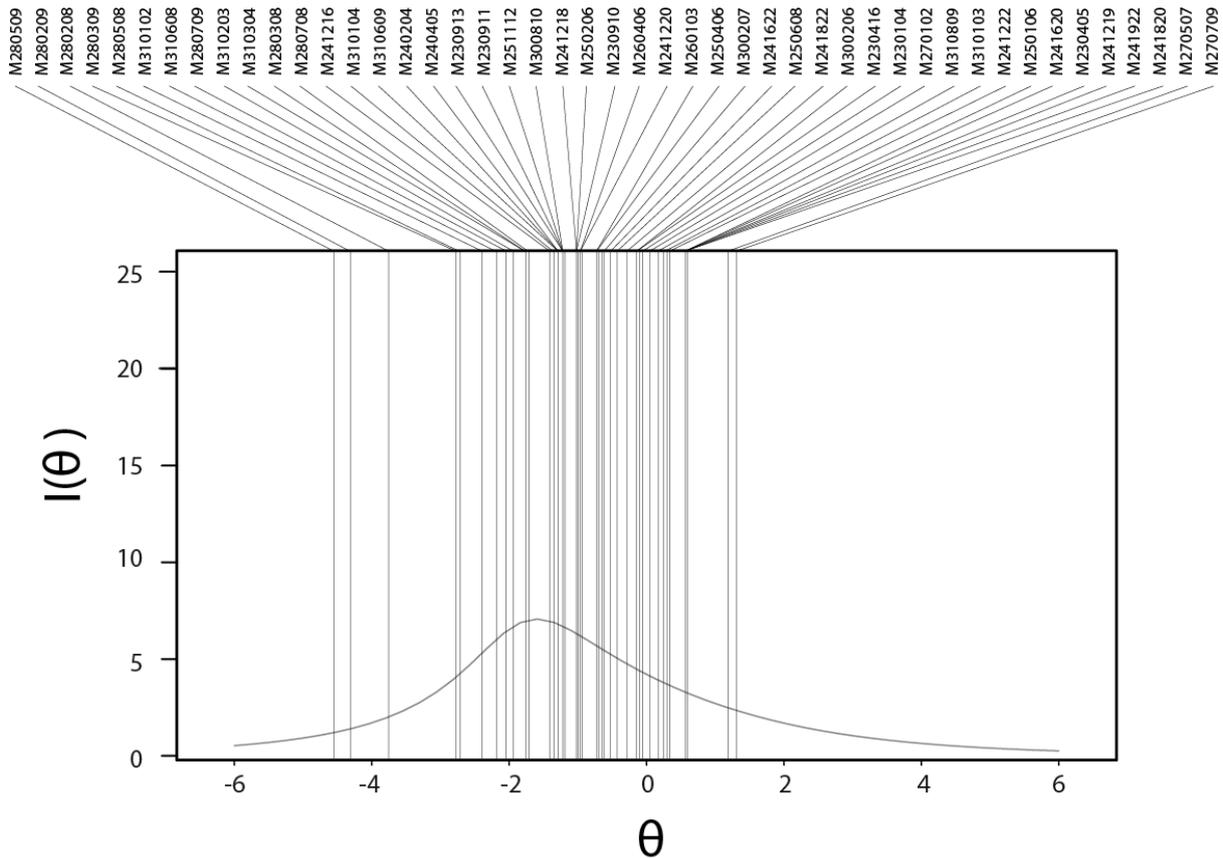


Abbildung 23: Testschwierigkeit des Chemie vignettentests (Datensatz: $NA < 50 = 0$) in Abhängigkeit von der Itemschwierigkeit, berechnet mit dem R-Package mirt (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. M steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung.

Bei den Chemiedatensätzen finden sich wie schon bei den Biologiedatensätzen kaum Unterschiede hinsichtlich der Testschwierigkeit. Wegen eines fehlenden Schwellenwertes ist von einer ungefähr gleichen Testschwierigkeit auszugehen. Große Unterschiede existieren nur hinsichtlich des Informationsgehaltes zwischen den Datensätzen $NA = 0$ und den beinahe gleichen Datensätzen $NA < 50 \% = 0$ und $NA < 50 \% = imp$.

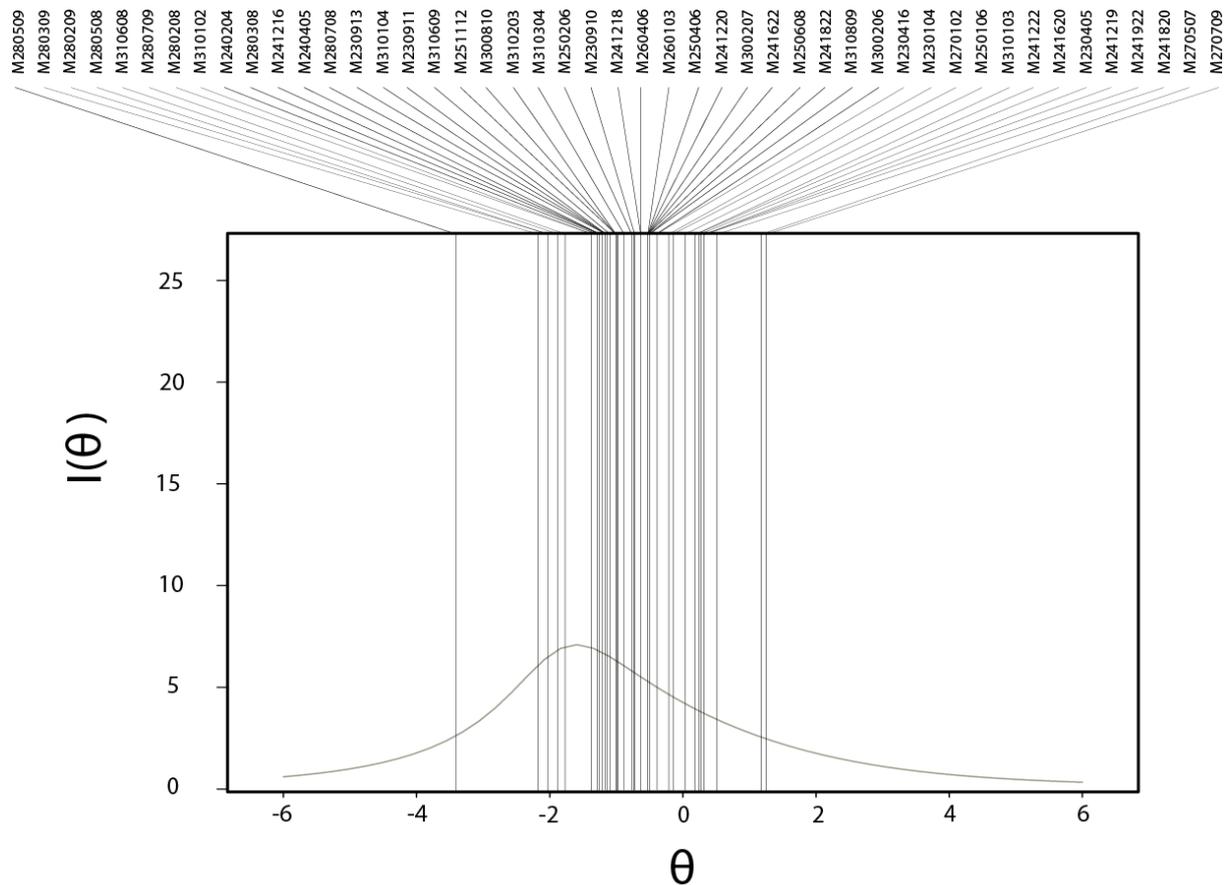


Abbildung 24: Testschwierigkeit des Chemie vignettentests (Datensatz: $NA < 50 = imp.$) in Abhängigkeit von der Itemschwierigkeit, berechnet mit dem R-Package `mirt` (Chalmers, 2014). θ gilt dabei als Maß der Personenfähigkeit in Abhängigkeit von der Itemschwierigkeit (vgl. Edelen & Reeve, 2007). $I(\theta)$ bezeichnet die Information bzgl. der Präzision und Reliabilität des Tests. M steht für Paarvergleich. Die ersten zwei Ziffern stehen für die Vignette, die letzten vier Ziffern für die beiden Items, die verglichen werden, eigene Darstellung.

Wiederum kann davon ausgegangen werden, dass die Hypothese H3-II („Die Testschwierigkeit liegt unabhängig vom Umgang mit fehlenden Werten im mittleren Schwierigkeitsbereich.“) *nicht falsifiziert* werden muss.

Sowohl für den Biologie- als auch für den Chemietest kann die mittlere Testschwierigkeit als weiterer Validitätshinweis verstanden werden. Dies rekurriert auch auf die Hypothesen H3-I und H3-II („Die Hauptaussagen der Ergebnisse aus dem Biologie- und Chemietest finden sich unabhängig vom Umgang mit fehlenden Werten wieder.“/„Die Testschwierigkeit liegt unabhängig vom Umgang mit fehlenden Werten im mittleren Schwierigkeitsbereich.“), da unabhängig vom Umgang mit fehlenden Werten immer von einem verhältnismäßig einfachen Test ausgegangen werden kann.

6.3 Empirische Überprüfung der Struktur des Noticings und des Reasonings zur Vermittlung von Modellkompetenz

In Kapitel 3.2.2 (S. 40) und 3.2.3 (S. 42) wurde erläutert, dass mit den entwickelten Vignetentests das Konstrukt *Professionelle Unterrichtswahrnehmung* (also Noticing und Reasoning) als spezielle Facette der Lehrkompetenz bei angehenden Lehrkräften untersucht wird. Bislang besteht Unklarheit darüber, was mit Noticing und Reasoning im Detail gemeint ist (Meschede, 2014, S. 18). Bezüglich möglicher Strukturen zur Vermittlung von Modellkompetenz im Biologie- und Chemieunterricht besteht bislang in der Literatur keine Einigkeit. Aus diesem Grund wurde sowohl zum Reasoning des Modelleinsatzes im naturwissenschaftlichen Unterricht als auch zum Noticing hinsichtlich der *Elemente zur Vermittlung von Modellkompetenz* eine eigene literaturgestützte Struktur entwickelt. Diese muss sich nun vor dem Hintergrund empirischer Daten beweisen (Modellierung). Hierzu werden IRT-basierte konfirmatorische Faktorenanalysen (CFA) vorgestellt, deren Kennwerte Aufschluss über die Passung zum theoretischen Modell geben.

6.3.1 Empirische Überprüfung der Struktur des Reasonings

Das hier verwendete Reasoning-Konstrukt umfasst beim Chemie- und Biologietest *Diagnosemöglichkeiten*, *Bewertungsmöglichkeiten* und *Handlungsalternativen auswählen können*. Es wurde überprüft, ob diese drei Facetten, die in der Formulierung der Fragestellung berücksichtigt wurden, auf die Struktur der Daten Einfluss nehmen. Auf diese Weise soll die Hypothese H1b-II („Die Passung der Reasoningbestandteile (*aus Diagnosemöglichkeiten*, *Handlungsmöglichkeiten* und *Unterrichtsbewertungen auswählen können*) lässt sich im Chemietest belegen.“) überprüft werden.

Im Chemiedatensatz muss H1b-II als *falsifiziert* gelten. Zum einen fielen sowohl *Diagnosemöglichkeiten*, *Bewertungsmöglichkeiten* als auch *Handlungsalternativen auswählen können* zusammen, zum anderen erreicht der nun entstandene Generalfaktor beim RMSEA zwar einen angemessenen Wert, CFI und TLI sind jedoch weit entfernt von den Anforderungen an einen guten Modellfit (vgl. Abbildung 25).

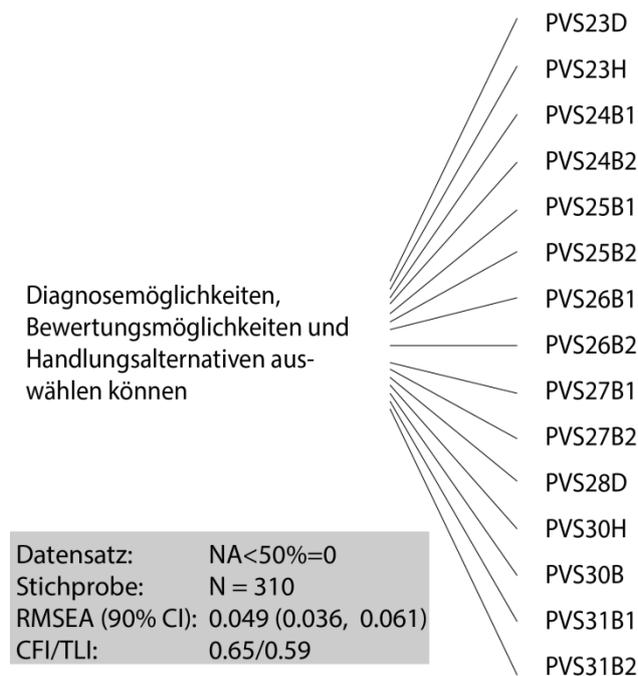


Abbildung 25: Reasoning CFA des Chemietests. Alle Itemscores fielen zu einem Generalfaktor zusammen. Ein fittendes Modell hinsichtlich RMSEA und CFI konnte über Fitindices nicht bestätigt werden. *Abkürzungen:* PVS: Paarvergleichssumme; B: Bewertungsmöglichkeiten auswählen können; D: Diagnosemöglichkeiten auswählen können; H: Handlungsalternativen auswählen können; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index, eigene Darstellung.

Auch der Chemiedatensatz $NA < 50\% = imp$ zeigt deutlich inakzeptable Werte, zumindest hinsichtlich CFI/TLI (CFI/TLI: 0.593/0.525; RMSEA: 0.051). Lediglich der Datensatz, bei dem alle fehlenden Werte mit 0 gewertet werden (Chemie $NA = 0$), zeigt gute Fitindices (CFI/TLI: 0.951/0.943; RMSEA: 0.051).

Ein Grund für das zumindest bei zwei Datensätzen nicht fittende Reasoningmodell kann im Ungleichgewicht der Paarvergleichssummen liegen: Es gab ein großes Übergewicht zugunsten des Bereichs *Bewertungsmöglichkeiten auswählen können* (elf Paarvergleichssummen; 73,3 %) während *Diagnosemöglichkeiten* (zwei Paarvergleichssummen; 11,3 %) und *Handlungsalternativen auswählen können* (zwei Paarvergleichssummen; 11,3 %) kaum ins Gewicht fielen. Zudem scheint der Einfluss der Fragestellung weniger entscheidend zu sein als der des Plots bzw. des Vignettenstammes.

In der Auswertung des Biologietests wurden die NoS-Items beim Reasoning nicht berücksichtigt, da diese das Wissen über Modellunterscheidungen abfragen und somit andere Itemfragestellungen umfassen. Auch hier fallen verschiedene Aspekte (*Diagnosemöglichkeiten und Handlungsalternativen*: Kovarianz bzw. latente Korrelation > 1) zusammen. Es verbleibt ein fittendes, zweifaktorielles Modell, das Abbildung 26 zu entnehmen ist.

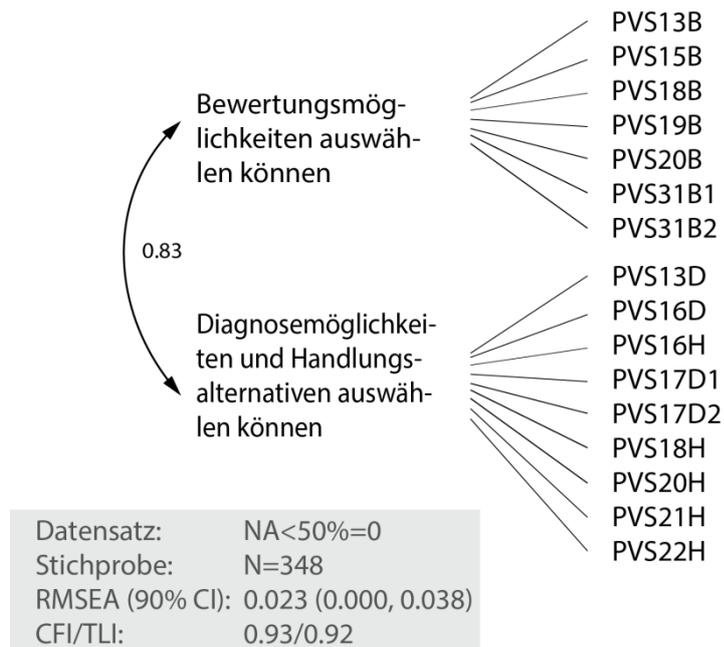


Abbildung 26: Reasoning CFA des Biologietests. Ein fittendes Modell hinsichtlich RMSEA und CFI wurde über Fitindices bestätigt. Die starke Korrelation zwischen Diagnosemöglichkeiten/Handlungsalternativen und Bewertungsmöglichkeiten ist signifikant. *Abkürzungen:* PVS: Paarvergleichssumme; B: Bewertungsmöglichkeiten auswählen können; D: Diagnosemöglichkeiten auswählen können; H: Handlungsalternativen auswählen können; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index, eigene Darstellung.

Dieses Ergebnis deckt sich auch mit den anderen Datensätzen. Der Datensatz $NA = 0$ erreicht auf diese Weise sehr gute Werte (CFI/TLI: 0.985/0.983; RMSEA: 0.028), der Datensatz $NA < 50\% = imp$ fittet akzeptabel (CFI/TLI: 0.917/0.904; RMSEA: 0.022).

Hypothese H1b-I („Die Passung der Reasoningbestandteile (aus *Diagnosemöglichkeiten*, *Handlungsmöglichkeiten* und *Unterrichtsbewertungen auswählen können*) lässt sich im Biologietest belegen.“) muss nur hinsichtlich der Facetten (Zusammenfall von *Diagnosemöglichkeiten* und *Handlungsalternativen auswählen können*) verändert, aber *nicht falsifiziert* werden. Die hohe Kovarianz bzw. latente Korrelation von 0.83 zwischen *Bewertungsmöglichkeiten auswählen können* und *Diagnosemöglichkeiten* sowie *Handlungsalternativen auswählen können* ähnelt beispielsweise den in Meschede (2014, S. 141) berichteten latenten Korrelationen zwischen den Reasoningfacetten *Veranschaulichung*, *Klarheit* und *Zusammenfassen und Hervorheben* (zwischen .80 und .93).

6.3.2 Empirische Überprüfung der Struktur des Noticings

Nach einem systematischen Literaturreview ergaben sich Strukturen dessen, was zur Vermittlung von Modellkompetenzen durch Lehrkräfte notwendig ist und demnach auch von Lehr-

amtsstudierenden und Lehramtsanwärterinnen und Lehramtsanwärttern wahrgenommen werden sollte. Trotz des geschlossenen Itemformats gehen wir davon aus, dass wir selektive Aufmerksamkeit messen, da die Hinweise durch geschlossene Items erst mit entsprechender Expertise wahrgenommen und in Form einer mit der Expertinnen- und Expertennorm übereinstimmenden Antwort verarbeitet werden können. Das Konstrukt Noticing (vgl. Kapitel 3.2.3, S. 42) wurde demnach sowohl für den Chemie- als auch für den Biologieunterricht entwickelt.

Hypothese H1a-II („Die Passung der aus einem systematischen Literaturreview gewonnenen Struktur des Noticings (*Elemente zur Vermittlung von Modellkompetenz*) lässt sich im Chemietest belegen.“) wird über die Berechnung der IRT-basierten konfirmatorischen Faktorenanalyse überprüft. Abbildung 27 fasst die Ergebnisse zusammen.

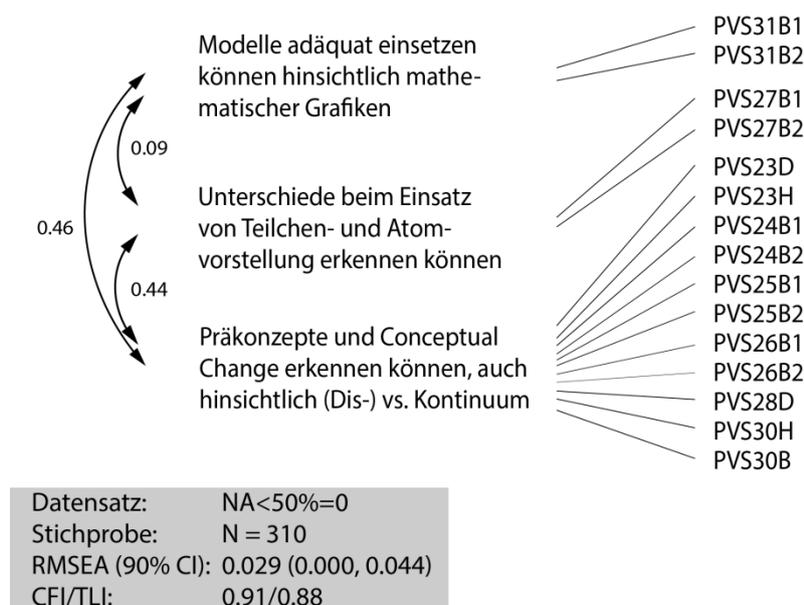


Abbildung 27: Modellkompeten-Noticing CFA Chemie. Ein fittendes Modell hinsichtlich RMSEA und CFI wurde gefunden, der TLI ist knapp nicht akzeptabel. *Abkürzungen:* PVS: Paarvergleichssumme; B: Bewertungsmöglichkeiten auswählen können; D: Diagnosemöglichkeiten auswählen können; H: Handlungsalternativen auswählen können; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index. Bis auf die Kovarianz zwischen *Modelle adäquat einsetzen können* und *Unterschiede beim Einsatz von Teilchen- und Atomvorstellungen erkennen können* ($p = 0.537$) sind die latenten Korrelationen signifikant, eigene Darstellung.

Die Modellierung des Chemiedatensatzes funktioniert bis auf den Fitindex TLI, der nur knapp unter dem Soll bleibt. Wird das in Abbildung 27 mit dem Imputationsdatensatz $NA < 50\% = imp$ gerechnet, so werden die Fitindizes schlechter und inakzeptabel (CFI/TLI: 0.802/0.761; RMSEA: 0.037). Wird vom Datensatz $NA = 0$ ausgegangen, so kann von einer

sehr guten Passung des theoretischen Modells des Noticings zu den empirischen Daten ausgegangen werden (CFI/TLI: 0.979/0.975; RMSEA: 0.034).

Für den Datensatz $NA < 50 \% = 0$ lag noch eine andere konfirmatorische Faktorenanalyse vor, bei der fünf Facetten (Faktor 1: *Präkonzepte*, Faktor 2: *Conceptual Change*, Faktor 3: *(Dis-)Kontinuum*, Faktor 4: *Atome vs. Teilchen*, Faktor 5: *Modelle adäquat einsetzen hinsichtlich mathematischen Grafiken*) unterschieden werden. Dieses fünfdimensionale Modell fittet zwar hervorragend (CFI/TLI: 0.995/0.993; RMSEA: 0.006), die Faktoren *Präkonzepte* und *(Dis-)Kontinuum* fallen jedoch zusammen (latente Korrelation > 1). Das daraus folgende Modell fittet schlechter.

Da der Datensatz $NA < 50 \% = 0$ als entscheidend definiert wurde, müsste die Hypothese H1a-II („Die Passung der aus einem systematischen Literaturreview gewonnenen Struktur des Noticings (*Elemente zur Vermittlung von Modellkompetenz*) lässt sich im Chemietest belegen.“) wegen des zu tiefen TLI-Werts *falsifiziert* werden, doch deuten die insgesamt doch befriedigenden Fit-Werte des Modells darauf hin, dass *eher keine Falsifizierung* angebracht ist.

Beim Biologietest wurde im Gegensatz zum Reasoning die NoS-Facette *Modelle vergleichen können* berücksichtigt. Somit entstand eine vierfaktorielle konfirmatorische Faktorenanalyse, die wiederum mit Mplus 7.3 (Muthén & Muthén, 2012) IRT-basiert (categorical; estimator = WLSMV) ausgewertet wurde. Dies dient der Überprüfung der Hypothese H1a-I („Die Passung der aus einem systematischen Literaturreview gewonnenen Struktur des Noticings (*Elemente zur Vermittlung von Modellkompetenz*) lässt sich im Biologietest belegen.“).

In Abbildung 28 ist die Struktur der konfirmatorischen Faktorenanalyse ersichtlich. Um Rücksicht auf die lokalstochastische Unabhängigkeit zu nehmen, wurden wiederum Paarvergleiche summiert und eine dieser Paarvergleichssummen wegen zu vieler Kategorien gruppiert (vgl. Kapitel 4.4, S. 68).

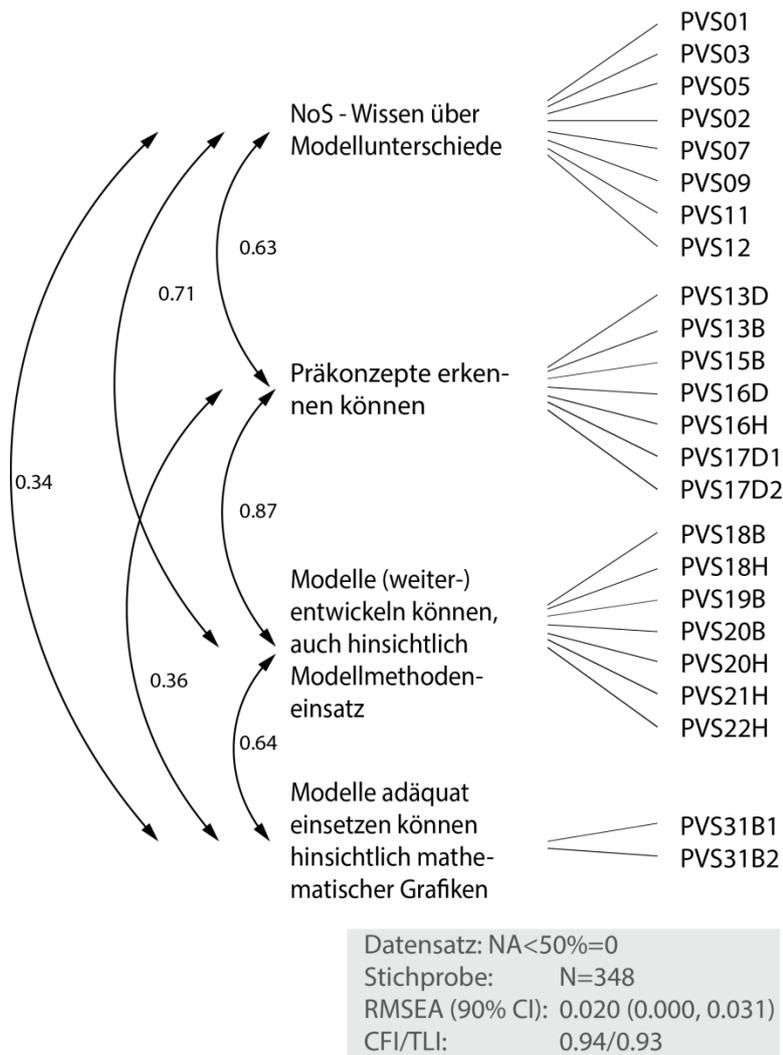


Abbildung 28: Noticing-CFA Biologie. Ein akzeptables bis gutes Modell hinsichtlich RMSEA sowie CFI und TLI wurde gefunden. Alle latenten Korrelationen werden signifikant. *Abkürzungen:* PVS: Paarvergleichssumme; B: Bewertungsmöglichkeiten auswählen können; D: Diagnosemöglichkeiten auswählen können; H: Handlungsalternativen auswählen können; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index, eigene Darstellung.

In Abbildung 28 wird ein fittendes Modell des Datensatzes $NA < 50\% = 0$ präsentiert. Der Datensatz $NA < 50\% = imp$ ergibt jedoch keine fittenden Werte (CFI/TLI: 0.877/0.862; RMSEA: 0.024). Der Datensatz $NA = 0$ fittet hingegen wieder sehr gut (CFI/TLI: 0.971/0.967; RMSEA: 0.033).

Die Kovarianz bzw. latente Korrelation zwischen *Präkonzepte erkennen können* und *Modelle (weiter-)entwickeln können* auch hinsichtlich der *Modellmethode* ist mit .87 hoch. Hohe Korrelationen bestehen ebenfalls zwischen *Modelle (weiter-)entwickeln können* auch hinsichtlich *Modellmethode* und dem *NoS-Wissen über Modellunterschiede* (.71) sowie zwischen *NoS* und *Präkonzepte erkennen können* (.63) und *Modelle (weiter-)entwickeln können* und *Modelle adäquat einsetzen können* hinsichtlich *mathematischer Grafiken* (.64). Deutlich ge-

ringer sind die Kovarianzen zwischen *Modelle adäquat einsetzen können* und *Präkonzepte erkennen können* (.36) und dem *NoS-Wissen* und *Modelle adäquat einsetzen können* (.34).

Die Hypothese H1a-I („Die Passung der aus einem systematischen Literaturreview gewonnenen Struktur des Noticings (*Elemente zur Vermittlung von Modellkompetenz*) lässt sich im Biologietest belegen.“) wird auf der Grundlage des grundlegenden Datensatzes $NA < 50 \% = 0$ nicht falsifiziert.

Zusammenfassend zeigt sich, dass sich trotz des aufwendigen systematischen Literaturreviews das theoretische Modell des Noticings (*Elemente zur Vermittlung von Modellkompetenz*) im Chemietest wegen eines knapp nicht akzeptablen TLI nur bedingt empirisch begründen lässt, zumal das Modell beim Datensatz $NA < 50 \% = 0$ nicht fittet; der Datensatz $NA = 0$ fittet aber gut. Hingegen fittet das theoretische Noticing-Modell des Biologietests empirisch, obwohl die Struktur mit vier Faktoren komplexer ist. Die latenten Korrelationen geben Aufschluss über Zusammenhänge, die hinsichtlich des Modelleinsatzes im Unterricht bisher noch unbekannt waren.

6.4 Exploratorische Faktorenanalysen

Konfirmatorische Faktorenanalysen haben das Ziel, theoretisch angenommene Strukturen auf ihre Passung zu empirischen Daten zu prüfen (Bortz & Döring, 2006, S. 517). Beim Chemietest zeigt sich eine nur bedingt hinreichende Passung des Noticing-Konstrukts. Bezüglich Reasoning ergibt sich hier eindeutig eine nicht passende Struktur.

Um Aussagen über die Struktur der Chemiedaten zu bekommen, wird hier ein explorativer Schritt eingefügt: die exploratorische Faktorenanalyse (EFA). Auf Grundlage der Paarvergleichssummen wurde mit Mplus 7.3 (Muthén & Muthén, 2012) eine IRT-basierte und mit Geomin rotierte exploratorische Faktorenanalyse mit dem Datensatz $NA < 50 \% = 0$ berechnet. Die Paarvergleichssumme MVS24B2 wurde umkodiert (der Wert 1 wurde in 0 umgewandelt), da sonst elf statt der für IRT vorausgesetzten Maximalzahl von zehn Kategorien (Mplus-Beschränkung) enthalten waren.

Die vierfaktorielle Version wurde auf Grundlage der Rotationsmethode Geomin Oblique signifikant als beste Version berechnet.²⁹ Die Passung zu den Daten kann als sehr gut bezeichnet

²⁹ Kommunalitäten (h^2) werden bei der Mplus-Ausgabe nicht angegeben.

werden (CFI/TLI: 0.998/0.996; RMSEA: 0.005). Die Faktorladungen sind Tabelle 15 zu entnehmen.

Tabelle 15: Mplus-basierte, geominrotierte Faktorladungen einer exploratorischen Faktorenanalyse mit dem Chemiedatensatz $NA < 50 \% = 0$

Faktor	1	2	3	4
MVS23D	0.383*	0.045	-0.213	-0.018
MVS23H	0.425*	-0.070	0.007	-0.062
MVS24B1	0.422*	0.085	-0.024	0.099
MVS24B2	-0.059	0.309*	0.117	0.086
MVS25B1	0.138	0.453*	0.027	0.025
MVS25B2	0.044	0.247	-0.309*	0.066
MVS26B1	0.245	0.134	-0.061	-0.123
MVS26B2	-0.065	0.173	-0.070	0.023
MVS27B1	-0.099	0.195	0.203	0.173
MVS27B2	0.005	0.388*	0.595*	-0.038
MVS28D	0.448*	-0.018	0.370*	0.279
MVS30H	0.026	0.597*	-0.023	-0.062
MVS30B	-0.060	0.348*	-0.040	0.191
MVS31B1	-0.011	-0.017	-0.178	0.649*
MVS31B2	0.089	0.039	0.011	0.653*

Anmerkungen: Die hier dargestellte vierfaktorielle Lösung fittet signifikant am besten; die Fitindices sind auch im Vergleich zu den konfirmatorischen Faktorenanalysen des Noticings und insbesondere des Reasonings deutlich besser.

* $p \leq .05$

Betrachtet man zur Interpretation nur die positiv signifikant ladenden Faktorladungen, so wird deutlich, dass Faktor 1, 2 und 3 schwierig zu interpretieren sind, insbesondere Faktor 2. So betrifft Faktor 1 zwar die Vignetten 23 und 24, welche zum Thema Präkonzepte konstruiert wurden und Vignette 28 zum Thema Conceptual Change, also zwei Themen, denen auch theoretisch eine nahe Verbindung zukommt. Allerdings war nicht zu erwarten, dass die zweite Itemfragestellung zum Bewerten der Unterrichtssituation MVS24B2 negativ auf diesen Faktor, jedoch auf Faktor 2 lädt. Dieser Faktor 2 wiederum hat außerdem signifikant Ladungen auf Fragestellung 1 zum Bewerten der Unterrichtssituation der Vignette 25 (MVS25B1), die im Kontext des Themas Diskontinuum und Kontinuum entwickelt wurde. Außerdem lädt die Fragestellung 2 (Bewerten) der Atome und Teilchen zugeordneten Vignette 27 (MVS27B2) sowie beide Fragestellungen der Vignette 30 signifikant, welche zu Conceptual Change entwickelt wurde. Faktor 3 hat ebenfalls bei der zweiten Fragestellung der Vignette 27 hohe, signifikante Ladungen, u. a. im Bereich Atome und Teilchen sowie dem theoretisch nicht unbedingt naheliegenden Conceptual Change (MVS28D). Am eindeutigsten ist Faktor 4, auf den Vignette 31 lädt. Diese Vignette beschäftigt sich mit der Passung mathematischer Darstellungen im Kontext Bildung für eine nachhaltige Entwicklung. Vignette 26 (MVS 26B1/B2) lädt auf keinen der Faktoren signifikant.

Zunächst wurde in Erwägung gezogen, dass das PID-Modell (Blömeke et al., 2015) die Faktoren interpretierbar machen könnte. So wurde überlegt, ob Faktor 1 und 2 als Decision making, Faktor 3 als Interpretation oder Decision making und Faktor 4 als Perceiving oder Interpretation bezeichnet werden könnte (Begriffe vgl. Santagata & Yeh, 2015, S. 154). Nach genauer Prüfung der Vignetten, Itemfragestellungen und Items musste diese Hypothese verworfen werden. Vielmehr scheint es sich bei Faktor 1 schlicht um den *fachdidaktischen Umgang mit Präkonzepten*, bei Faktor 2 um den *fachdidaktischen Umgang mit (Modell-)Vorstellungen* zum Übergang zwischen *Realmodellen* und *Modellen des Diskontinuums*, bei Faktor 3 evtl. um *Teilchenmodelle* bzw. nicht sprachliche Modelle und bei Faktor 4 um den *fachdidaktischen Umgang mit Grafiken* bzw. *gezeichneten Modellen* zu handeln.

Hier wird davon ausgegangen, dass das hinter dem PID-Modell (Blömeke et al., 2015) stehende Konstrukt beim Umgang mit den hier entwickelten Vignetten vermutlich Einfluss nimmt, dass dieses aber in der naturwissenschaftlichen Unterrichtsrealität bzw. der Wahrnehmung dieser Unterrichtsrealität von ggf. stärkeren Faktoren überlagert wird. Die fachdidaktische Kompetenz scheint zumindest beim Chemietest prägender auf die Wahrnehmung der Probandinnen und Probanden zu wirken als die Kompetenz im Umgang mit Vignetten.

Unter den gleichen statistischen Maßgaben wie beim Chemietest wurde beim Biologietest ebenfalls eine exploratorische Faktorenanalyse gerechnet. Obwohl die dreifaktorielle Version knapp signifikant besser als die zweifaktorielle Lösung ($p = 0.048$) berechnet wurde, musste auf die zweifaktorielle Lösung zurückgegriffen werden (RMSEA: 0.010; CFI/TLI: 0.987/0.984), da die Fitindices der dreifaktoriellen Lösung nichtpassende Ergebnisse (RMSEA: 0.000; CFI/TLI: ≥ 1) lieferten.

Tabelle 16: Mplus-basierte, geominrotierte Faktorladungen einer exploratorischen Faktorenanalyse mit dem Biologiedatensatz $NA < 50\% = 0$

Faktor	1	2
MNVS13D	0.503*	0.048
MNVS13B	0.316*	0.164*
MNVS15B	0.237*	-0.103
MNVS16D	0.355*	-0.031
MNVS16H	0.295*	-0.030
MNVS17D1	0.187*	-0.023
MNVS17D2	0.043	0.117
MNVS18B	0.313*	0.061
MNVS18H	0.198*	-0.087
MNVS19B	0.416*	0.003
MNVS20B	0.314*	0.250*
MNVS20H	0.342*	0.051
MNVS21H	0.495*	-0.071
MNVS22H	0.391*	0.060
MNVS31B1	0.413*	-0.062

MNVS31B2	0.347*	0.074
MNVS01	0.170*	0.061
MNVS03	0.247*	0.189*
MNVS05	0.395*	-0.062
MNVS02	0.492*	-0.064
MNVS07	0.344*	0.461*
MNVS09	-0.002	1.136*
MNVS11	0.124	0.116
MNVS12	0.302*	0.051

Anmerkungen: Die zweifaktorielle Lösung fittet zwar signifikant schlechter als die dreifaktorielle Lösung, bei der dreifaktoriellen entstehen jedoch nichtmögliche Fitindices (CFI/TLI: ≥ 1). Die zweifaktorielle Lösung ergibt zwar bessere Fitindices als die konfirmatorischen Faktorenanalysen zum Noticing und Reasoning, jedoch ist der Abstand deutlich geringer als beim Chemiedatensatz.

* $p \leq .05$

Dieses Ergebnis ist auf Grundlage der Planung des Tests nicht interpretierbar. Weder zeichnet sich eine klare Unterscheidung von Noticing und Reasoning ab, noch verbinden sich Vignetten in einer anderen Weise, die auf der Basis anderer Theorien verständlich wäre. Da die Unterschiede der Fitindices zu den konfirmatorischen Faktorenanalysen zu Noticing (CFI/TLI: 0.941/0.934; RMSEA: 0.020) und Reasoning (CFI/TLI: 0.930/0.919; RMSEA: 0.023) nicht groß sind, kann für den Biologietest der Bedarf einer explorativ gewonnenen Struktur zurückgewiesen und die konfirmatorische Struktur beibehalten werden.

7 Diskussion, Limitation, Ausblick und Zusammenfassung

Das Schlusskapitel hat folgenden Aufbau: Zunächst werden fachdidaktisch relevante, dann statistische Ergebnisse diskutiert. Im Anschluss werden die Limitationen der Studie erörtert sowie ein Ausblick auf mögliche anschließende Forschung abgeleitet. Die wichtigsten Ergebnisse werden schließlich vor dem Hintergrund der Forschungshypothesen (Kapitel 3.4, S. 57) zusammengefasst.

7.1 Diskussion

Empirisch erhobene Daten über die Häufigkeit von bestimmten Unterrichtsphasen im alltäglichen Unterricht sind verhältnismäßig selten. Ein Grund dafür könnte sein, dass Unterrichtsbeobachtungen angekündigt sind. Es ist anzunehmen, dass durch die Ankündigung der Unterricht von der Lehrkraft „inszeniert“ (vgl. Henkenborg, Krieger, Pinseler & Behrens, 2008, S. 41) wird und damit abweicht vom üblichen Unterricht. Aussagen über tatsächliche Häufigkeiten bestimmter Unterrichtsphasen stehen deshalb im Zweifel.

Quantifizierbare und belastbare empirische Aussagen über die Häufigkeit des Modelleinsatzes im naturwissenschaftlichen Unterricht liegen folglich nicht vor. Dennoch ist davon auszugehen, dass der Modelleinsatz einen verhältnismäßig hohen Anteil an der Unterrichtszeit einnimmt, mitunter, weil der Mensch als „modellbildendes Wesen ... *an* und *mit* Modellen“ lernt und „*nach* Modellen“ (Stachowiak, 1980, S. 9) handelt.

Umso mehr überrascht es, dass ein Schwerpunkt fachdidaktischer Forschung auf anderen Arbeitsweisen wie beispielsweise dem Experimentieren liegt. Auch scheint der Vermittlung von Modellkompetenz weniger Bedeutung in der praktischen Lehramtsausbildung eingeräumt zu werden als anderen Inhalten.³⁰ Eventuell wird von der impliziten Annahme ausgegangen, dass Vermittlung von Modellkompetenz eine Lehrkompetenz ist, die nicht erlernt zu werden braucht. Allerdings sprechen die Befunde zur ungenügenden Modellkompetenz von Schülerinnen und Schülern (z. B. Günther, Fleige, Upmeier zu Belzen & Krüger, 2017; Krell et al., 2014) ebenso gegen diese Annahme wie die Befunde zum Erfolg von Seminaren für künftige Lehrkräfte zur Vermittlung von Modellkompetenz (Günther et al., 2017; J. H. van Driel, O. de Jong & N. Verloop, 2002; van Driel & Jong, 2015; van Driel & Verloop, 1999, 2002).

³⁰ Ggf. gilt dies noch mehr für die Biologie- als die Chemieausbildung.

All dies deutet darauf hin, dass der Vermittlung von Modellkompetenz bei der Lehramtsausbildung mehr Aufmerksamkeit eingeräumt und ihre Anforderungen stärker in den Blick genommen werden sollte. Dies müsste möglichst konkret geschehen (vgl. van Driel & Verloop, 2002), um für die Praxis relevant zu sein.

Studien zur Modellkompetenz von Schülerinnen und Schülern (z. B. Krell et al., 2014) und theoretische Erkenntnisse (z. B. Beerenwinkel & Parchmann, 2008) bestehen zwar, bislang fehlt es jedoch an einer Übersicht, die professionell fundierte *Elemente zur Vermittlung von Modellkompetenz* benennt. Diese Übersicht steht nun als Ergebnis dieser Studie durch ein systematisches Literaturreview zur Verfügung (vgl. Abbildung 4 und Abbildung 5). Es wäre jetzt möglich, die *Elemente zur Vermittlung von Modellkompetenz* ins Curriculum der Lehrkräfteausbildung aufzunehmen. Durch die in Klammern angegebenen Häufigkeiten der Nennungen der *Elemente zur Vermittlung von Modellkompetenz* in den Artikeln, lassen sich ungefähre Hinweise auf die Dringlichkeit der Vermittlung an künftige Lehrkräfte ableiten.

Außerdem eignen sich die *Elemente zur Vermittlung von Modellkompetenz* als Grundlage zur Operationalisierung von Lehrkompetenztests in diesem Bereich. Bei der Entwicklung der hier vorgestellten Tests dieser Studie konnten nicht alle *Elemente zur Vermittlung von Modellkompetenz* berücksichtigt werden (vgl. Kapitel 7.2, S. 115). Wünschenswert wäre die Erfassung aller *Elemente zur Vermittlung von Modellkompetenz*, um umfassende Aussagen zur Qualität der Vermittlung machen zu können. Dies kann, wie hier geschehen, als Test zur Leistungserhebung (künftiger) Lehrkräfte dienen. Später wäre ein Einsatz zum Assessment in doppelter Hinsicht sinnvoll: Zum einen könnte so über die Eignung fürs Lehramt hinsichtlich der Vermittlungsfähigkeit zur Modellkompetenz entschieden werden. Dies scheint jedoch nur dann sinnvoll zu sein, wenn weiter davon ausgegangen wird, dass Vermittlung von Modellkompetenz nicht erlernbar ist. Demnach wäre es zum anderen wahrscheinlich wichtiger und für die Praxis relevanter, mithilfe eines passenden Testinstruments eine differenzierte Diagnose von Fähigkeiten und Problemen bei bereits zugelassenen zukünftigen und bereits tätigen Lehrkräften zu ermöglichen. Ein solches, nahe an der Unterrichtswirklichkeit entwickeltes Testinstrument stellen die hier entwickelten Vignettentests (s. u.) zur Verfügung. Für zukünftige Lehrkräfte könnten Seminare auf der Grundlage der hier vorgelegten Übersicht über die *Elemente zur Vermittlung von Modellkompetenz* individuelle Problemelemente thematisieren und so die Fähigkeiten der einzelnen Lehrkraft individuell wirkungsvoll fördern (Günther et al., 2017; Jong et al., 2005a; Jong & van Driel, 2004; Justi & van Driel, 2005a, 2005b). Entsprechendes könnte für Fortbildungen bereits erfahrener Lehrkräfte geschehen. Vielversprechende Ansätze zur wirkungsvollen Förderung sind insbesondere in der Präsentation und Re-

flexion der entstandenen Videovignetten und weiterer Unterrichtsvideos zur Vermittlung von Modellkompetenz zu sehen (vgl. den Ansatz des Videoclubs bei Sherin, 2007).

Sowohl die über das systematische Literaturreview erstellte Übersicht über die Modelldefinitionen als auch die Noticing-Strukturdiagramme zu den *Elementen zur Vermittlung von Modellkompetenz* zeigen zwischen den Unterrichtsfächern Chemie und Biologie große Unterschiede. Zwar ist im Schulfach Biologie von einer größeren Diversität hinsichtlich der Modelldefinitionen auszugehen, weil die Biologie im Vergleich zur Chemie und Physik mit der „Mannigfaltigkeit ... der belebten Natur“ (Peters, 1998, S. 102) umzugehen hat. Im Schulfach Chemie zeigt sich jedoch eine größere Bandbreite hinsichtlich der *Elemente zur Vermittlung von Modellkompetenz*. Dies liegt sicher auch daran, dass „[d]ie Einführung submikroskopischer Modellvorstellungen (...) ein neuralgischer Punkt für das Verständnis und Interesse für Chemie“ (Saborowski, 2000, S. 13) ist und der Conceptual Change von Präkonzepten hier tendenziell jäher und sprunghafter erfolgen muss als beispielsweise in Biologie und Physik (vgl. Saborowski, 2000). Infolgedessen besteht in der Fachdidaktik Chemie eine Diskussion darüber, auf welche Weise der Conceptual Change im Schulfach Chemie am besten erfolgen kann.

Um näher an der Unterrichtswirklichkeit zu operieren, wurde in dieser Studie je ein Vignettest zur Vermittlung von Modellkompetenz im Biologie- und Chemieunterricht entwickelt. Bei Vignettestests werden im Gegensatz zur Unterrichtsbeobachtung stellvertretend Unterrichtssituationen in Text- und Videoform präsentiert und in dieser Studie auf einer sechsstufigen Skala zwischen „trifft gar nicht zu“ bis „trifft voll und ganz zu“ bewertet. Kersting et al. (2012) berichten von Zusammenhängen zwischen dem Testscore eines Vignettestests und beobachteter Unterrichtsdurchführung sowie dem Lernerfolg von Schülerinnen und Schülern. Ein Nachweis dieser Zusammenhänge wird auch für die in dieser Studie entstandenen Vignettestests angestrebt (vgl. Kapitel 7.3, S. 116).

Das mit Vignettestests erhobene Konstrukt wird unterschiedlich benannt und steht in der Diskussion (vgl. Kapitel 3.2, S. 38). In der hier vorgestellten Studie wird vom Konstrukt *Professionelle Unterrichtswahrnehmung* (Goodwin, 1994; Sherin, 2007; Sherin & van Es, 2009; Steffensky et al., 2015; Stürmer & Seidel, 2015) ausgegangen. Weitgehende Einigkeit besteht inzwischen, dass mit *Professioneller Unterrichtswahrnehmung* selektive Aufmerksamkeit (*Noticing*) und wissensbasiertes Schlussfolgern (*Reasoning*) gemeint ist. Diskutiert wird, wie *Noticing* und *Reasoning* definiert werden können (vgl. Meschede, 2014, S. 18). Uneinigkeit besteht bereits beim *Reasoning*, noch stärker jedoch beim *Noticing*. Im Gegensatz zu Ansät-

zen, welche versuchen, insbesondere das *Noticing*, aber auch das *Reasoning* allgemeingültig zu definieren (Stürmer & Seidel, 2015, S. 55), gehen wir davon aus, dass sich fachdidaktische Professionalität bei der Unterrichtswahrnehmung durch eine Adaption an unterschiedliche Unterrichtssituationen zeigt. Daraus ergibt sich die Notwendigkeit einer Definition des *Noticings* und *Reasonings* speziell in Hinblick auf die Vermittlung von Modellkompetenz. In Anlehnung an bisherige Definitionen des *Reasonings* (Kleinknecht, 2014, S. 7; Kleinknecht & Schneider, 2013, S. 14; Seidel & Stürmer, 2014, S. 745; Sherin, 2007, S. 390; Stürmer & Seidel, 2015, S. 55) wird hier *Reasoning* mit der *Fähigkeit, aus Diagnosemöglichkeiten, Unterrichtsbewertungen und Handlungsalternativen auswählen zu können*, definiert.

Beim *Noticing* geht es um die Frage nach der Situation oder dem Inhalt, der oder dem aus einer professionellen Perspektive selektive Aufmerksamkeit geschenkt werden sollte. Deshalb ist es hier notwendig, abweichend von bisherigen Definitionen (vgl. z. B. Stürmer & Seidel, 2015, S. 55) *Noticing* nicht als eng umrissene allgemeinpädagogische Fähigkeit, sondern als *Elemente zur Vermittlung von Modellkompetenz* zu definieren (vgl. Abbildung 4 und Abbildung 5). Die Definitionselemente von *Noticing* wurden aus dem erwähnten systematischen Literaturreview abgeleitet. *Noticing* kann laut Meschede et al. (2015, S. 320) über geschlossene Items nicht erhoben werden, da durch solche Items bereits Aufmerksamkeitsfokuse vorgegeben würden. In der hier vorgestellten Studie wird hingegen davon ausgegangen, dass die Studierenden erst mit zunehmender Expertise anfangen, die Hinweise durch die Items wahrzunehmen und darauf ihren Aufmerksamkeitsfokus zu verändern. Es wird also von einer Expertiseschwelle ausgegangen. Für diese These spricht der beim Biologie- und Chemietest signifikante Testscorezuwachs querschnittlich über die gruppierten Semester und beim Chemietest zusätzlich der signifikante Testscorezuwachs über die geleisteten und gruppierten Unterrichtsstunden. Daraus abgeleitet nehmen Novizinnen und Novizen offensichtlich die in den Items enthaltenen Hinweise auf Fokuse der Aufmerksamkeit nicht wahr. Gegen das Argument des Vorliegens einer Expertiseschwelle spricht die verhältnismäßige Einfachheit der Tests: Die IRT-skalierte durchschnittliche Testschwierigkeit liegt beim Biologie- und Chemietest bei ca. $\theta = -0.5$.³¹ Daraus kann abgeleitet werden, dass die Tests erst bei schwierigeren Items gut zwischen Expertinnen bzw. Experten und Novizinnen bzw. Novizen zu trennen vermögen sowie den Abstufungen (gruppierte Semesteranzahl bzw. Unterrichtsstunden) dazwischen. Die Expertiseschwelle scheint also eher im Bereich der Hälfte der schwierigeren

³¹ Der Test trennt also zwischen der querschnittlichen Semesteranzahl und zumindest in Chemie der querschnittlichen Anzahl geleisteter Unterrichtsstunden. Aber insgesamt ist der Test verhältnismäßig leicht.

Items zu liegen. Um die Testzeit zu reduzieren, erscheint eine Reduktion der Itemanzahl im Bereich der leichteren Hälfte der Items hilfreich zu sein (vgl. Kapitel 7.3, S. 116).

Sowohl das theoretisch abgeleitete Konstrukt des Noticing als auch des Reasonings wurde über das Verfahren der konfirmatorischen Faktorenanalyse, die Konstruktstrukturen auf ihre Passung getestet, überprüft. Angestrebt wurde also, Hinweise über die Passung dieser Konstrukte zu den mit den Vignettentests erhobenen empirischen Daten zu erhalten. Beim Biologievignettentest kann laut der Fitindizes RMSEA, CFI und TLI sowohl von einer Passung des Konstrukts Noticing (*Elemente zur Vermittlung von Modellkompetenz*) als auch des Konstrukts Reasoning (*Fähigkeit, aus Diagnosemöglichkeiten, Unterrichtsbewertungen und Handlungsalternativen auswählen zu können*) ausgegangen werden. Damit wird erstmals ein Vignettentest zur *Vermittlung von Modellkompetenz* vorgelegt, welcher sowohl Noticing als auch Reasoning abdeckt, also Professionelle Unterrichtswahrnehmung als Ganzes.

Die Annahme, dass das Reasoning-Konstrukt zu den empirischen Daten des Chemie-vignettentests passt, muss falsifiziert werden. Gemäß der Fitindizes RMSEA und CFI kann jedoch von einer Passung des Noticing-Konstrukts zu den Daten ausgegangen werden; nur der TLI liegt zwei Hundertstel unterhalb des geforderten Schwellenwerts (Hu & Bentler, 1998, S. 449). Andere Studien berichten den TLI nicht (Krauss et al., 2011, S. 148), obwohl bei CFI = 1.00 und RMSEA = .00 ein TLI > 1 möglich wäre (vgl. Hu & Bentler, 1998, 436 f.). Konfirmatorische Faktorenanalysen mit ähnlichen Fitindex-Werten wurden in der vorliegenden Studie wegen eines TLI > 1 nicht berichtet.

Im Folgenden werden die latenten Korrelationen bzw. Kovarianzen von Elementen zur Vermittlung von Modellkompetenz diskutiert. Beim Chemie-vignettentest (vgl. Abbildung 27, S. 99) zeigen die empirischen Daten ein Zusammenfallen der Elemente *Präkonzepte* und *Conceptual Change*. Dies entspricht der fachdidaktischen Theorie. Außerdem lässt sich beobachten, dass die zum Element (*Dis-*) vs. *Kontinuum* entwickelte Vignette mit *Präkonzepte* und *Conceptual Change* zusammenfällt. Das wurde so nicht erwartet. Bei genauerer Betrachtung kann dieses Phänomen jedoch damit begründet werden, dass sich Chemiedidaktik zentral mit der Überwindung von Präkonzepten zum diskontinuierlichen Atomverständnis beschäftigt. Daraus ergibt sich eine dreifaktorielle konfirmatorische Faktorenanalyse zu den Elementen zur Vermittlung von Modellkompetenz im Chemieunterricht mit folgenden Faktoren:

1. *Unterschiede beim Einsatz von Teilchen und Atomvorstellungen erkennen können*
2. *Modelle adäquat einsetzen können hinsichtlich mathematischer Grafiken*
3. *Präkonzepte und Conceptual Change erkennen können, auch hinsichtlich (Dis-) vs. Kontinuum.*

Bei der konfirmatorischen Faktorenanalyse sind die Elemente *Unterschiede beim Einsatz von Teilchen und Atomvorstellungen erkennen können* und *Modelle adäquat einsetzen können hinsichtlich mathematischer Grafiken* latent klar getrennte Konstrukte (.09). Dies entspricht der Schulerfahrung und gibt wiederum Hinweise auf die Validität des Chemievignettentests. *Modelle adäquat einsetzen können hinsichtlich mathematischer Grafiken* und *Präkonzepte und Conceptual Change erkennen können, auch hinsichtlich (Dis-) vs. Kontinuum* korrelieren erwartungskonform moderat (.46). Ungefähr in gleicher Höhe korrelieren erwartungskonform *Präkonzepte und Conceptual Change erkennen können, auch hinsichtlich (Dis-) vs. Kontinuum* sowie *Unterschiede beim Einsatz von Teilchen und Atomvorstellungen erkennen können* (.44).

Beim Biologievignettentest (vgl. Abbildung 28, S. 101) wird das *NoS-Wissen über Modellunterscheidung* ebenso wie *Präkonzepte erkennen können* als eigenständiges Konstruktelement abgebildet. *Modelle adäquat einsetzen können hinsichtlich mathematischer Grafiken* bleibt ebenfalls ein eigenständiges Element. Das Element *Modellweiterentwicklung* fällt mit dem Element *Modellmethode* zusammen, was vor dem Hintergrund der Schulpraxis nachvollziehbar ist. Daraus ergibt sich eine vierfaktorielle konfirmatorische Faktorenanalyse zu den Elementen zur Vermittlung von Modellkompetenz im Biologieunterricht mit folgenden Faktoren:

1. *NoS-Wissen über Modellunterscheidung*
2. *Präkonzepte erkennen können*
3. *Modelle (weiter-)entwickeln können, auch hinsichtlich Modellmethodeneinsatz*
4. *Modelle adäquat einsetzen können hinsichtlich mathematischer Grafiken.*

Erwartet wurden moderate bis hohe Korrelationen zwischen den Elementen *Modelle adäquat einsetzen können hinsichtlich mathematischer Grafiken* und dem *NoS-Wissen über Modellunterscheidung*. Die Erwartung erfüllte sich nicht (.34). Möglicherweise geht der eher fachdidaktisch-methodische *adäquate Modelleinsatz hinsichtlich mathematischer Grafiken* über das fachwissenschaftliche *NoS-Wissen über Modellunterscheidung* hinaus. Das Strukturelement *Modelle (weiter-)entwickeln können auch hinsichtlich Modellmethodeneinsatz* hängt erwartungskonform stark mit *Präkonzepte erkennen können* (.87) und mit dem *NoS-Wissen über Modellunterschiede* (.71) zusammen. Die Strukturelemente *Modelle (weiter-)entwickeln können auch hinsichtlich Modellmethodeneinsatz* und *Modelle adäquate einsetzen können hinsichtlich mathematischer Grafiken* hängen erwartungskonform eng zusammen (.64). Dass *NoS-Wissen über Modellunterschiede* und *Präkonzepte erkennen können* so eng latent korrelieren (.63) unterstützt die Befunde, dass Fachwissen auch fachdidaktisch relevant ist. Kontra-

intuitiv ist die geringe latente Korrelation von *Präkonzepte erkennen können* und *Modelle adäquat einsetzen können hinsichtlich mathematischer Grafiken* (.36). Dies kann mit der starken Verankerung von *Präkonzepte erkennen können* in der Fachdidaktik begründet werden, während *Modelle adäquat einsetzen können hinsichtlich mathematischer Grafiken* unterrichtsplanerische und -methodische Aspekte betrifft.

In Abbildung 29 sind die drei zentralen Ergebnisse zusammengefasst, die auf die Wirksamkeit der Lehramtsausbildung hinsichtlich der Fähigkeit zur Vermittlung von Modellkompetenz Einfluss haben dürften.

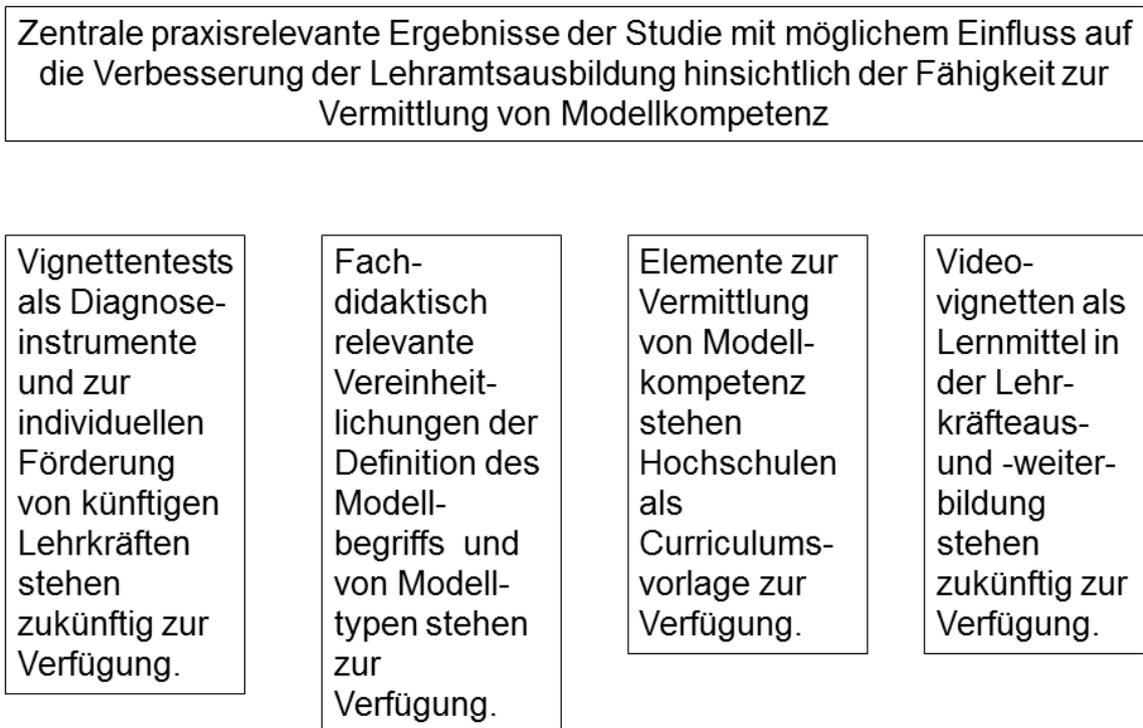


Abbildung 29: Studienergebnisse mit besonderer Relevanz für die Unterrichtspraxis, eigene Darstellung.

Abbildung 29 fasst bedeutsame Ergebnisse der Studie für die Unterrichtspraxis zusammen. Dies wird im Folgenden ergänzt um die Diskussion statistischer und testtheoretischer Aspekte, u. a. zur Validität des Tests. Wie sich in Kapitel 6.2 (S. 79) zeigt, gibt es sowohl beim Chemie- als auch beim Biologietest Hinweise darauf, dass es sich um kriteriums- und inhaltsvalide Tests handelt. Aufgrund der Rückmeldungen durch Expertinnen- und Expertenbefragungen ist außerdem von ökologischer Generalisierbarkeit auszugehen.

Die in dieser Studie entstandenen Tests mussten sich mit der Forderung nach lokalstochastischer Unabhängigkeit auseinandersetzen. Hierzu gilt es, folgendes Kriterium zu beachten: „Wann ist die Annahme der lokalen stochastischen Unabhängigkeit verletzt? Auf jeden Fall bei logischen Abhängigkeiten zwischen den Items. Logische Unabhängigkeit zwischen den Items bedeutet, daß die Beantwortung eines Items nicht eine bestimmte Antwort auf ein anderes Item voraussetzen darf.“ (Rost, 2004, 74 f.) Positions- und Reihenfolgeeffekte sind zumindest bei der Onlineumsetzung der Tests durch die Rotation der Items und Vignetten im Onlinetool SoSci-Survey (Leiner & Leiner, 2015) weitgehend ausgeschlossen. Meschede et al. (2015, S. 327) schlagen zum Umgang mit der lokalstochastischen Unabhängigkeit das Nested-Factor-Modell vor, d. h. die Vignettenitems lassen sich einerseits den Vignetten, andererseits einem Generalfaktor in einer konfirmatorischen Faktorenanalyse zuordnen. Der Chemiedatensatz war für solch ein Vorgehen zu komplex und konvergiert nicht. Beim Biologiedatensatz wurde dies zwar entsprechend berechnet, aufgrund einer Fehlermeldung fehlen jedoch entscheidende Angaben, um das Ergebnis zu berichten. Aus diesem Grund fiel in dieser Studie die Entscheidung auf ein anderes Vorgehen: Die Itemwerte wurden vor dem Hintergrund der Fragestellung zur jeweiligen Itematterie summiert. Für eine IRT-Skalierung wäre das Nested-Factor-Modell (vgl. Kapitel 7.2, S. 115) zwar wünschenswert gewesen; um lokalstochastische Unabhängigkeit zu gewährleisten, ist dieses aber nicht zwingend erforderlich (vgl. Rost, 1996). Allerdings sind diesbezüglich weitere, elaboriertere Vorgehen geplant.

Da es hier um die Bestimmung von Strukturen im Zusammenhang mit der Vermittlung von Modellkompetenzen an Schülerinnen und Schüler geht, wird die Frage nach der nicht ausreichenden Passung der empirischen Chemiedaten zum theoretischen Modell diskutiert. Generell, das heißt sowohl für das Reasoning als auch für das Noticing, könnte die Stichprobe zu klein sein. Systematische, künstliche Ergänzungen, beispielsweise durch künstliche Verdopplung der Stichprobe, brachten keine Verbesserung (vgl. Kapitel 5.3, S. 72). Dennoch fällt auf, dass die fittende Biologie-Stichprobe mit $N = 348$ größer war als die nichtfittende Chemiestichprobe mit $N = 310$. Da der Fitindex TLI beim Chemie-Noticing nur knapp unter den geforderten 0.9 bleibt, führt eine geringfügige Ergänzung, beispielsweise durch Daten von Lehrkräften im Vorbereitungsdienst, evtl. zu einem ausreichenden Fit, da – wie im Kapitel 6.2.2 (S. 84) gezeigt wurde – diese zumindest beim Biologietest einen besseren Gesamtscore (bzw. beim Chemietest keinen signifikant schlechteren Gesamtscore) als im siebten bis achten Semester aufweisen. Diese Gruppe war beim Chemietest jedoch besonders unterrepräsentiert ($n = 11$).

Bezüglich der knapp nicht fittenden konfirmatorischen Faktorenanalyse zum Noticing im Chemietest ist auch die große Heterogenität der Stichprobe zu berücksichtigen. Im Gegensatz zu anderen EKoL-Teilprojekten, die hauptsächlich in baden-württembergischen Pädagogischen Hochschulen Daten erhoben haben, wurden in dieser Studie Probandinnen und Probanden von Hochschulen in Norddeutschland bis in die Schweiz befragt. Damit sollte in den Naturwissenschaften, insbesondere im Lehramt Chemie, eine ausreichende Stichprobe sichergestellt werden. Es ist somit von einer größeren (auch kulturellen) Diversität auszugehen (vgl. Brovelli, Kauertz, Rehm & Wilhelm, 2011; Sunder, Todorova & Möller, 2015) als bei den anderen Teilprojekten. Umso erstaunlicher ist die gute Passung des Modells des Biologietests. Eventuell unterscheiden sich die Ausbildungsgänge und -inhalte für das Lehramt in der Chemie stärker von Hochschule zu Hochschule als in der Biologie, was zu einer größeren Heterogenität der Daten und damit komplexeren Datenstrukturen führen könnte.

Zudem mussten zur Sicherstellung einer ausreichenden Stichprobengröße verschiedene Erhebungsmodi angeboten werden (Internetversion im Seminar und von zu Hause aus, beim Biologietest zusätzlich noch eine Paper-Pencil-Variante). Insbesondere das Ausfüllen im Internet von zu Hause aus führte im Biologietest post hoc zu höchst signifikanten Unterschieden im Vergleich zur Paper-Pencil-Version im Seminar hinsichtlich komplett fehlender Werte bei Vignetten. Beim Chemietest wird der Unterschied zwischen der Internetversion im Seminar und von zu Hause hochsignifikant hinsichtlich der komplett fehlenden Werte. Probandinnen und Probanden, die von zu Hause aus den Vignettentest ausfüllen, arbeiten demnach mitunter weniger gründlich. Dies führt zu einer Verfälschung der empirischen Daten, was z. B. mit Aufgabenverweigerung oder dem Phänomen des Abbruchs bei Onlineerhebungen begründet werden könnte (vgl. Baur & Florian, 2009, 122 ff.).

Mit den fehlenden Werten kann unterschiedlich umgegangen werden. Eine in der Studie bislang nicht berücksichtigte Möglichkeit des Umgangs mit fehlenden Werten ergibt sich aus der IRT-Schätzung. Allerdings besteht hier das Problem der lokalstochastischen Unabhängigkeit (vgl. Kapitel 4.4, S. 68). Aus diesem Grund wurden Paarvergleichssummen eingesetzt. Allerdings ergibt eine IRT-basierte Schätzung auf dieser Grundlage keinen Sinn, da dabei Paarvergleichssummen je Fall mit fehlenden Werten geschätzt würden, was ungenaue Schätzungen entstehen lassen würde. Eine andere Möglichkeit, um Hinweise auf die Bedeutung der fehlenden Werte in den Datensätzen zu erhalten und so evtl. Rückschlüsse auf adäquate Imputationsverfahren erhalten zu können, könnte das Bootstrapping sein.

Zuletzt ist auf die systematische Aufarbeitung des Umgangs mit fehlenden Werten hinzuweisen (vgl. als Überblick: Lüdtke et al., 2007). Der Datensatz $NA = 0$ hätte sowohl beim Chemie- als auch beim Biologietest sehr gute Fitwerte erzeugt. Dass damit allen Personen, die Items nicht beantwortet haben, „Unfähigkeit“ unterstellt wird ($= 0$), ist problematisch, aber nicht völlig auszuschließen. Zwar generiert dieser Datensatz sowohl beim Biologie- als auch beim Chemietest die besten Fitwerte. Dass es sich um ein Artefakt handelt (Hypothese: Konfirmatorische Faktorenanalysen fitten immer besser, wenn man Probandinnen und Probanden mit vielen Nullen ergänzt), konnte jedoch durch eine Simulation mit einem Vergleichsdatensatz zurückgewiesen werden. Die multiple Imputation auf Grundlage aller Items, Kovariaten und Demografie erbrachte durchgehend die schlechtesten Fitwerte. Dass der Datensatz $NA = 0$ so gute Werte erbringt, sollte insbesondere beim Chemietest Veranlassung für weitere Erhebungen mit den entstandenen Vignettentests sein. Damit könnte ein Trend hin zur Passung oder Nichtpassung des theoretischen Modells zu den empirischen Daten geliefert werden.

7.2 Limitation

Modelle und deren Einsatz im Unterricht hängen stark davon ab, welches Objekt oder Phänomen betrachtet wird. Daher ist eine Repräsentation verschiedener Modelle, die im Unterricht häufig eingesetzt werden, im Vignettentest in einer möglichst großen Bandbreite wünschenswert, allerdings aufgrund der Unendlichkeit der denkbaren Modelle und Themen quasi unmöglich. Über das systematische Review konnten zwar Elemente von Modellkompetenzen, die Lehrkräfte bei der Modellvermittlung haben sollten, extrahiert werden. Eine adäquate Passung zu den in der Schule verwendeten Modellen konnte jedoch nicht sichergestellt werden. Eine möglichst repräsentative Auswahl von typischen Modellen im Unterricht wäre wünschenswert; insbesondere beim Chemietest wäre eine Erweiterung der Themen angebracht. Auch beim Biologietest sind aufgrund der Fülle nicht alle typischen Modelle berücksichtigt; es gelang jedoch m. E., eine größere Breite der Themen abzudecken. Außerdem konnten in den entwickelten Vignettentests nicht alle *Elemente zur Vermittlung von Modellkompetenz* berücksichtigt werden (vgl. Kapitel 7.1, S. 106).

Ähnlich zeigt sich die Diversität kontextabhängiger Themen zum Modelleinsatz auch in den empirischen Daten. Während beim Biologiedatensatz eine zweidimensionale exploratorische Faktorenanalyse am besten fittete und auch ein Generalfaktormodell akzeptable Fitindices ergab, wurde beim Chemiedatensatz eine vierfaktorielle Version als die am signifikant besten

passende bestimmt. Die Interpretation und Benennung, insbesondere des zweiten Faktors, ist schwierig. Es scheint sich um eine Mischung aus Noticing- und Reasoning-Facetten zu handeln, die evtl. über die Wahrnehmung von Vignetteninhalt und Frage verständlich wird. Insofern böte sich eine Think-aloud-Studie an. Zum einen zur Detektion der Wahrnehmung der Probandinnen und Probanden, zum anderen zur diesbezüglichen Modifizierung der Vignetten. Dies schliesse bspw. an die Arbeiten von Roberts, Stewart und Pullin (2006) an, die die Entwicklung des Denkens über Modelle zur Genetik über Think-aloud-Protokolle analysierten. Auszuschließen ist jedoch nicht, dass man beim Chemietest im Gegensatz zum Biologietest evtl. besser nicht mehr von Professioneller Unterrichtswahrnehmung ausgehen sollte, sondern von dem Konstrukt Topic-Specific Pedagogical Content Knowledge (vgl. Aydin, Friedrichsen, Boz & Hanuscin, 2014) *Modellkompetenz im Chemieunterricht vermitteln können*. Dieses könnte an das in Kapitel 2.1.2 (S. 21) dargestellte Kompetenzmodell von COACTIV (Baumert & Kunter, 2011) anschließen.

Bislang wurden insbesondere im Chemie- aber auch beim Biologietest, zu wenige Lehrkräfte im Vorbereitungsdienst befragt, um Hinweise auf die These von der Expertiseschwelle bei geschlossenen Items und damit auf die Frage zu erhalten, ob über geschlossene Items Noticing erhoben werden kann (vgl. Kapitel 7.1, S. 106).

Um mit einer der Voraussetzungen für die IRT-Skalierung, also der lokalstochastischen Unabhängigkeit, umzugehen, wäre das von Meschede (2014) vorgeschlagene Nested-Factor-Modell ein Gewinn. Allerdings ist die Datenstruktur insbesondere des Chemietests dafür zu komplex und deshalb aufgrund von beschränkter PC-Leistung nicht möglich. Da über Paarvergleichssummen beim Reasoning nicht alle Bestandteile lokalstochastischer Unabhängigkeit erfüllt sind, nur der Einfluss der Fragestellung zu den Itembatterien berücksichtigt wird und nicht immer die Vignettenebene, ist eine Vereinfachung der Struktur des Chemie-vignettestests unter Vorbehalt erstrebenswert.

7.3 Ausblick

Der Ausblick geht zunächst auf generell ergänzungswürdige Aspekte, sodann auf Validierungsgesichtspunkte und im Anschluss auf statistische Alternativen ein.

Eine Einigung auf eine didaktisch fundierte Modelldefinition und einen Modellbegriff, zumindest für das jeweilige Unterrichtsfach Chemie und Biologie, wäre wünschenswert. Denn die große Bandbreite der Begriffe führt momentan dazu, dass zwar die gleichen Worte ver-

wendet, aber unterschiedliche Vorstellungen und Kontexte gemeint werden. Einheitliche Modelldefinitionen auch der Modelltypen könnten dazu führen, dass diese im Lehramtsstudium spezifischer vermittelt werden würden und in ihrer Umsetzung im Unterricht lernförderlich wirken könnten. Im Umkehrschluss würde das hinter dem Vignettentest stehende Modellkonstrukt einheitlicher verstanden und der Test somit wahrscheinlich reliabler bzw. die Fitindices der konfirmatorischen Faktorenanalysen besser. Denn die Probandinnen und Probanden bringen aufgrund ihrer unterschiedlichen Vorstellungen zu Modellen zusätzlich eine Varietät in die Antworten, der nur schwer begegnet werden kann. Die Vermittlung einer einheitlichen Modelldefinition als Teil des Curriculums der Lehramtsausbildung wäre folglich hilfreich, um durch eine solche Einheitlichkeit der Modelldefinitionen auch Klarheit für Schülerinnen und Schüler zu schaffen. In Anbetracht der vorgestellten Ergebnisse scheint eine Schulung der Professionellen Unterrichtswahrnehmung zu *Elementen zur Vermittlung von Modellkompetenz* schon zu Beginn des Lehramtsstudiums sinnvoll.

Zur Überprüfung der Stimmigkeit der Annahme, dass auch geschlossene Items Noticing erheben können, wäre ein Vergleich hilfreich. Sinnvoll wäre eine Substudie, welche als Think-aloud-Studie angelegt sein könnte (vgl. Kapitel 7.2, S. 115). Diese könnte folgendermaßen gestaltet sein: Künftigen Lehrkräften wird zunächst nur die Vignette vorgelegt. Sie äußern dazu ihre Gedanken, lesen im Anschluss die Items und äußern dann, was sie sich hierzu denken. Anschließend beantworten sie die Items. Transkripte der Äußerungen ließen dann Rückschlüsse darauf zu, ob sich durch die geschlossenen Items der Aufmerksamkeitsfokus in Richtung Expertenmeinung verschiebt.

Bei der Testentwicklung ist die Frage der Validität entscheidend. Erstrebenswert wäre hierfür die Überprüfung der Korrelation der Testleistung mit Unterrichtsqualität und dem Lernerfolg der Schülerinnen und Schüler (vgl. Kersting et al., 2012). Solche Überprüfungen haben im Rahmen dieser Studie nicht stattgefunden. Bezüglich der Modellkompetenz der Schülerinnen und Schüler liegen kontextunabhängige Tests für den Chemie- und Biologieunterricht von Krell et al. (2014) vor, die im Längsschnitt mehrebenenanalytische Zusammenhänge zum Vignettentest während des Schulpraktikums oder Referendariats erlauben würden. Interessant wäre diesbezüglich auch die Überprüfung der Korrelation der Testleistung und Unterrichtsqualität bei erfahrenen Lehrkräften, um weitere belastbare Hinweise auf die Validität der Tests zu erhalten.

Zuvor sollte jedoch der Chemietest dahingehend weiterentwickelt werden, dass weniger Komplexität hinsichtlich der Dimensionen besteht. Drei Möglichkeiten werden hierzu disku-

tiert: erstens die Reduktion der Testzeit durch Reduktion der Itemanzahl im Bereich der leichteren Hälfte der Items (vgl. Kapitel 7.1, S. 106), da der Test tendenziell zu leicht ist und die Expertiseschwelle erst bei der schwereren Hälfte der Items zu trennen scheint. Zweitens wäre eine Verkleinerung hinsichtlich der Noticing-Strukturelemente denkbar. Davon wird aus fachdidaktischen Gründen eher abgeraten, da Hinweise auf die komplexe Struktur der *Elemente zur Vermittlung von Modellkompetenz* hilfreich für die Lehramtsausbildung wären. Drittens wäre im Umkehrschluss die Vergrößerung der Vignettenanzahl in den bisher unterrepräsentierten Vignetten *Modelle adäquat einsetzen können hinsichtlich mathematischer Modelle* und insbesondere im Bereich *Atome- vs. Teilchenvorstellungen vermitteln* sinnvoll. Ähnliches wäre ebenfalls wünschenswert für das Reasoning. Das Nichtfitten der dreidimensionalen Struktur des Reasoning-Konstrukts (Zusammenfall von *Diagnosemöglichkeiten* und *Handlungsalternativen auswählen können*) hängt beim Biologietest möglicherweise auch mit dem Ungleichgewicht der Anzahl der Fragen zu den Itembatterien *Bewerten, Diagnose* und *Handlungsalternativen* zusammen. Ein Angleichen der Itemanzahl wäre also sinnvoll (vgl. Geiser, 2011), auch um Aussagen über Schwierigkeitsunterschiede beim *Diagnostizieren, Bewerten* und *Handlungsalternativen auswählen* zu treffen. Beim Chemietest trifft dies noch stärker zu. Eine Ergänzung von *Diagnose-* (z. B. zur Heterogenität in der Klasse) und *Handlungsalternative-*Items wäre sinnvoll.

In diesem Zuge sollte auf die Komplexität der Chemiedaten reagiert werden und wie schon beim Biologiedatensatz eine Ergänzung durch die Erhebung der NoS-Aspekte *Wissen über Modellunterscheidungen* erfolgen. Auf den ersten Blick erscheint dieser Vorschlag kontraintuitiv. Jedoch ist es dadurch möglich, falls sich die Struktur weiterhin als besonders komplex herausstellen sollte, wichtige Rückschlüsse auf den Ursprung der Vermittlung von Modellkompetenz zu ziehen und daraufhin komplexitätsreduzierende Maßnahmen zu ergreifen.

Statistisch wäre eine Überprüfung interessant, ob sich die berichteten manifesten Korrelationen zu Kovariaten auch bei latenter Berechnung ähnlich zeigen würden, um tatsächlich Hinweise auf Unterschiede von Largescale-Studien und Vignettentests hinsichtlich beispielsweise der Korrelation von Fachdidaktik und Fachwissen zu erhalten (vgl. Voss et al., 2011). Für Rückschlüsse auf die Korrektheit verschiedener Datensätze, bei denen fehlende Werte unterschiedlich imputiert wurden, könnte das Bootstrapping wichtige Hinweise liefern, da so der Einfluss der Imputationsart auf die Fitindizes detailliert nachvollzogen werden könnte.

7.4 Zusammenfassung

In dieser Studie wurde je ein Vignettentest in den Schulfächern Chemie und Biologie entwickelt; beide Vignettentests beschäftigten sich mit der Vermittlung von Modellkompetenz. Ziel war es, einen validen und reliablen Test zu entwickeln, der diesbezügliche Lehrkompetenz indirekt über die professionelle Unterrichtswahrnehmung erhebt. Andere Studien zeigen, dass Ergebnisse aus Vignettentests Prädiktoren für Lehrqualität und Leistungszuwachs von Schülerinnen und Schülern sein können (Kersting et al., 2012). Ob auch die hier entstandenen Vignettentests solch eine Vorhersage für den konkreten Unterricht ermöglichen, konnte im Rahmen dieser Studie noch nicht erfasst werden.

Dem Ziel der Entwicklung eines validen Tests folgend wurde zunächst (Lehr-)Kompetenz definiert. Um die Schwierigkeiten des Anspruchs auf die Erhebung von Lehrkompetenz deutlich zu machen, wurden anschließend bisherige, auf deklarativem Wissen basierende Paper-Pencil-Tests vorgestellt und deren Ergebnisse kritisch untersucht.

Hierauf folgte die Darstellung eines anderen Testformats, dessen Anspruch es ist, näher an der Unterrichtswirklichkeit und damit auch an der Lehrkompetenz zu sein: der Vignettentest. Es galt zu beantworten, was ein Test misst, der Unterrichtssituationen präsentiert. In dieser Studie wird davon ausgegangen, dass weder direkt Lehrkompetenz noch allein (deklaratives) Wissen erfasst wird, sondern *Professionelle Unterrichtswahrnehmung*. Dieses aus selektiver Aufmerksamkeit (Noticing) und wissensbasiertem Schlussfolgern (Reasoning) bestehende Konstrukt kann mit der Lehrkompetenz zusammenhängen (vgl. Kersting et al., 2012), ist jedoch nicht mit dieser identisch.

Dies geschieht vor dem Hintergrund einer Darstellung von den unterschiedlichen Modelltypen in der Schule, der Definition von Modellkompetenz und deren Bedeutung im Biologie- und Chemieunterricht. Da es in den Tests darum geht, diejenigen Aspekte von Unterricht wahrnehmen zu können, die für das Vermitteln von Modellkompetenz relevant sind, wird hierzu der Terminus Noticing benutzt. Im Gegensatz zu Meschede et al. (2015) wird in dieser Studie davon ausgegangen, dass trotz der Präsentation von geschlossenen Items selektive Aufmerksamkeit bzw. Noticing erhoben wird. Der Grund dafür ist, dass Novizinnen und Novizen die in den Items enthaltenen Hinweise für das Ausrichten der Aufmerksamkeit noch nicht wahrnehmen können. Der in Kapitel 6.2.2 (S. 84) aufgezeigte querschnittlich signifikante Anstieg der Vignettenscores über die Anzahl der gruppierten Fachsemester unterstreicht diese Annahme.

Es folgte eine Darstellung der entwickelten Methoden, insbesondere der Entwicklung der Vignetten, der damit verbundenen Validierungsschritte und der Bepunktung des Tests. Die anschließende Ergebnisdarstellung lieferte Hinweise zur Validität. Zunächst wurde die akzeptable bis gute Reliabilität der Tests aufgezeigt, dann wurden aus erwartungskonform signifikanten Korrelationen zu Kovariaten Validitätshinweise gezogen.

Wichtiges Ergebnis ist die moderate Korrelation der generellen, nicht modellspezifischen Fachwissenstests Chemie und Biologie mit den Scores des Biologie- und Chemie vignettentests. Obwohl zur Vermittlung von Modellkompetenz ein hohes Fachwissen angenommen wird, manifestiert sich im Gegensatz zu herkömmlichen, deskriptives Wissen abfragenden Largescale-Studien kein annähernder Zusammenfall von Fachwissen und Fachdidaktik bei latenten Korrelationen von .91 (Voss et al., 2011). Die hier erhobene fachdidaktische Kompetenzfacette zur Vermittlung von Modellkompetenz ist eigenständig. Der Chemie vignettentest korreliert mit dem chemischen Fachwissenstest mit $r = .32$, der Biologie vignettentest mit dem biologischen Fachwissenstest mit $r = .41$.

Beim Chemie vignettentest zeigt sich eine mittlere ($r = .39$), beim Biologie vignettentest eine geringe Korrelation ($r = .25$) zum pädagogischen Unterrichtswissen (König & Blömeke, 2010). Dass bei Vignettentests auch pädagogisches Wissen Einfluss nimmt, war anzunehmen. Angenommen wurde jedoch eine höhere Korrelation zum Fachwissen als zum pädagogischen Unterrichtswissen. Dies war beim Chemie vignettentest nicht der Fall. Der Biologie vignettentest war diesbezüglich erwartungskonform.

Die dritthöchste (und dennoch niedrige) Korrelation ergab sich zwischen den Testscores und dem Wilde-Intelligenz-Test (Kersting et al., 2008). Dieser dient neben der Überprüfung der Intelligenz zugleich als Maß des Zusammenhangs von Textvignetten mit sprachlichen Fähigkeiten. Ziel war der Ausschluss, dass Textvignetten allein über sprachliche Intelligenz gelöst werden können. Mit Korrelationen von $r_{\text{Chemie}} = .33$ und $r_{\text{Biologie}} = .34$ besteht zwar ein Zusammenhang, aber nur ein geringer, was als weiterer Validitätshinweis interpretierbar ist.

Die weiteren, nur geringen Korrelationen sind Kapitel 6.2.1 (S. 80) zu entnehmen. Für die zukünftige Forschung wäre eine Überprüfung, ob sich die manifesten Korrelationen auch bei latenter Berechnung zeigen, wünschenswert.

Es folgte ein Kapitel zu Zusammenhängen zwischen Testscores und soziodemografischen Gesichtspunkten. Hierbei zeigt sich außer beim NoS-Wissenstest_{Biologie} *Modelle unterscheiden können* ($r = .19$) erwartungskonform kein signifikanter Zusammenhang zwischen dem Alter

der Probandinnen und Probanden und den Testscores. Erwartungskonform war jedoch die Korrelation zwischen Abiturnote und Biologiegesamtttestscore ($r = -.23$)³²; der Chemietestscore korreliert dagegen nicht signifikant mit der Abiturnote. Ebenso hat das Geschlecht keinen Einfluss. Beim Biologietest finden sich zwischen Studiengang und Testscore post hoc keine signifikanten Unterschiede. Beim Chemietest gibt es jedoch marginal signifikante Unterschiede zwischen Studierenden des Gymnasiallehramts und Personen, die kein Lehramt studierten.

Wichtige Validitätshinweise ergeben sich aus Zusammenhängen zur unterrichtspraktischen Erfahrung und der querschnittlich erhobenen Anzahl der Fachsemester. So zeigt sich beim Chemietest post hoc ein signifikanter Unterschied bei den Werten *keine bis acht Stunden* und *mehr als 50 Stunden* Unterrichtserfahrung. Dies entsprach der Erwartung und ist damit ein weiterer Hinweis auf die Kriteriumsvalidität des Tests. Beim Biologietest lassen sich diese Unterschiede nicht nachweisen.

Hinsichtlich der Anzahl der Fachsemester zeigen sich beim Chemietest querschnittlich marginal signifikante Anstiege zwischen erstem bis drittem vs. viertem bis sechstem Semester. Außerdem zeigen sich signifikante Zuwächse zwischen Studienanfängerinnen und Studienanfängern und Studierenden im siebten bis achten Semester. Darauf folgt ein nicht signifikanter Abfall hin zum Referendariat. Beim Biologietest steigt der durchschnittliche Testscore des ersten bis dritten zum siebten bis achten Semester signifikant sowie zwischen Studienbeginn und Referendariat hochsignifikant, was für beide Tests als Hinweis auf Kriteriumsvalidität gewertet wird. In Anbetracht dieser Ergebnisse scheint eine Schulung der Professionellen Unterrichtswahrnehmung schon zu Beginn des Lehramtsstudiums sinnvoll.

Inhaltsvalidität kann auch mit dem Nachweis begründet werden, dass unabhängig vom Umgang mit fehlenden Werten von einer ähnlichen Testschwierigkeit ausgegangen werden kann. Diese konnte im Kapitel 6.2.3 (S. 89) gezeigt werden.

Wichtiges Ziel der Studie war die Modellierung des Noticings, dessen theoretische Struktur sich aus einem systematischen Literaturreview ergab und das ein Desiderat darstellt. Mit dem als Standard gesetzten Datensatz $NA < 50 \% = 0$ konnten beim Chemietest zwar nicht alle, aber die meisten geforderten Fitindices erreicht werden (CFI/TLI: 0.905/0.886; RMSEA: 0.028). Bei anderem Umgang mit fehlenden Werten, z. B. mit dem Datensatz $NA = 0$ wurden

³² Die Korrelation mit negativem Vorzeichen ergibt sich dadurch, dass beim deutschen Notensystem 1 die beste und 6 die schlechteste Schulnote ist.

alle geforderten Fit-Werte erfüllt (CFI/TLI: 0.979/0.975; RMSEA: 0.034), beim Imputationsdatensatz $NA < 50 \% = imp$ hingegen nicht (CFI/TLI: 0.802/0.761; RMSEA: 0.037). Es ist folglich von einer bedingten Passung des theoretisch angenommenen Modells zu den empirischen Daten auszugehen.

Beim Biologietest passt die theoretische Struktur beim entscheidenden Datensatz $NA < 50 \% = 0$ (CFI/TLI: 0.941/0.934; RMSEA: 0.020). Gründe hierfür werden im Kapitel 6.3.1 (S. 96) diskutiert. Es ist hiermit von einer Passung des theoretisch angenommenen Modells zu den empirischen Daten auszugehen.

Als Reasoning bzw. Fähigkeit zum wissensbasierten Schlussfolgern wurden in dieser Studie *Diagnosemöglichkeiten*, *Bewertungsmöglichkeiten* und *Handlungsalternativen auswählen können* definiert. Diese drei Facetten wurden in den Fragestellungen zu den Items aufgegriffen. Über konfirmatorische Faktorenanalysen wurde überprüft, ob sich diese Aufteilung auch in den empirischen Daten wiederfindet. Für den Chemietest kann diese Annahme klar zurückgewiesen werden. Alle drei Facetten fallen hier zusammen. Das entstehende Generalfaktormodell fittet bei dem Standarddatensatz $NA < 50 \% = 0$ nicht (CFI/TLI: 0.647/0.588; RMSEA: 0.049). Gleiches gilt für den Datensatz $NA < 50 \% = imp$ (CFI/TLI: 0.593/0.525; RMSEA: 0.051). Davon abweichend fittet $NA = 0$ gut (CFI/TLI: 0.951/0.943; RMSEA: 0.051).

Beim Biologiedatensatz $NA < 50 \% = 0$ fallen *Diagnosemöglichkeiten* und *Handlungsalternativen auswählen können* zusammen. *Bewertungsmöglichkeiten auswählen können* bleibt eigenständig. Das hieraus gewonnene Modell fittet akzeptabel (CFI/TLI: 0.930/0.919; RMSEA: 0.023).

Beim erneuten Einsatz des Chemie- wie auch Biologietests sollten weitere Itembatterien zu den unterrepräsentierten Dimensionen des Reasonings *Diagnosemöglichkeiten* und *Handlungsalternativen auswählen können* entwickelt werden.

Die Ergebnisse zu Validierung und Modellierung sind in Tabelle 17 zusammengefasst.

Tabelle 17: Überblick über die Passung verschiedener konfirmatorischer Faktorenanalysen und Validitätshinweise in Abhängigkeit von den Datensätzen

Fit indices		Validierung			
CFI/TLI	RMSEA	Lehrstunden	Fachsemester	Alter/Geschlecht/Erhebungsstandort	Sonstiges/Anmerkung
Chemie-Noticing: $NA = 0$					
0.979/ 0.975	0.034	Insgesamt: m.s. Post-hoc- Untergruppen:	1.-3. vs. 7.-8 FS: m.s.. überall leichter n.s. Abfall zum	n.s./n.s./n.s.	PrKCCDIS- KONT, ATTEI, MAE:

		n.s.	Ref.		latent getrennt
Chemie-Noticing: $NA < 50\% = 0$					
0.905/ 0.886	0.028	0-8 h vs. mehr als 50 h: sig.	1.-3. vs. 4.-6. FS: m.s. 1.-3. vs. 7.-8 FS: sig.	n.s./n.s./n.s.	PrKCCDIS- KONT, ATTEI, MAE: latent getrennt
<hr/>					
0.995/ 0.993	0.006	0-8 h vs. mehr als 50 h: sig.	1.-3. vs. 4.-6. FS: m.s. 1.-3. vs. 7.-8 FS: sig.	n.s./n.s./n.s.	PrK+DISKONT (fallen zusam- men), CC, ATTEI, MAE
Chemie-Noticing: $NA < 50\% = imp$					
0.802/ 0.761	0.037 0.010		0-8 h vs. mehr als 50 h: sig.	1.-3. vs. 4.-6. FS: m.s. 1.-3. vs. 7.-8 FS: sig.	PrKCCDIS- KONT, ATTEI, MAE: latent getrennt
Chemie-Reasoning: $NA = 0$: D, H und B fallen zusammen zu GF					
0.951/ 0.943	0.051	Insgesamt: m.s. Post-hoc- Untergruppen: n.s.	1.-3. vs. 7.-8 FS: m.s.. leichter n.s. Abfall zum Ref.	n.s./n.s./n.s.	GF
Chemie-Reasoning: $NA < 50\% = 0$: D, H und B fallen zusammen zu GF					
0.647/ 0.588	0.049	0-8 h vs. mehr als 50 h: sig.	1.-3. vs. 4.-6. FS: m.s. 1.-3. vs. 7.-8 FS: sig.	n.s./n.s./n.s.	GF
Chemie-Reasoning: $NA < 50\% = imp$: D, H und B fallen zusammen zu GF					
0.593/ 0.525	0.051	0-8 h vs. mehr als 50 h: sig.	1.-3. vs. 4.-6. FS: m.s. 1.-3. vs. 7.-8 FS: sig.	n.s./n.s./n.s.	GF
<hr/>					
Biologie-Noticing: $NA = 0$:					
0.971/ 0.967	0.033	n.s.	1.-3. vs. 7.-8 FS: m.s. 1.-3. FS vs. Ref.: sig 4.-6. FS vs. Ref: m.s.	n.s./n.s./n.s.	PrK, NoS, MWEMM, MAE
Biologie-Noticing: $NA < 50\% = 0$:					
0.941/ 0.934	0.020	n.s.	1. FS vs. Ref. = sig 1.-3. vs. 7.-8 FS: sig 1.-3. FS vs. Ref: hochsig.	n.s./n.s./n.s.	PrK, NoS, MWEMM, MAE
Biologie-Noticing: $NA < 50\% = imp$:					
0.877/ 0.862	0.024	n.s.	1. FS vs. Ref. = sig 1.-3. vs. 7.-8 FS: sig 1.-3. FS vs. Ref: hochsig.	n.s./n.s./n.s.	PrK, NoS, MWEMM, MAE
Biologie-Reasoning: $NA = 0$: D und H fallen zusammen					
0.985/ 0.983	0.028	n.s.	1.-3. vs. 7.-8 FS: m.s. 1.-3. FS vs. Ref.: sig 4.-6. FS vs. Ref: m.s.	n.s./n.s./n.s.	DH, B
Biologie-Reasoning: $NA < 50\% = 0$: D und H fallen zusammen					
0.930/ 0.919	0.023	n.s.	1. FS vs. Ref. = sig 1.-3. vs. 7.-8 FS: sig 1.-3. FS vs. Ref: hochsig.	n.s./n.s./n.s.	DH, B
Biologie-Reasoning: $NA < 50\% = imp$: D und H fallen zusammen					
0.917/ 0.904	0.022	n.s.	1. FS vs. Ref. = sig 1.-3. vs. 7.-8 FS: sig 1.-3. FS vs. Ref: hochsig.	n.s./n.s./n.s.	DH, B

Abkürzungen: m.s. = marginal signifikant; n.s. = nicht signifikant; sig. = signifikant; PrK = Präkonzepte; CC = Conceptual Change; ATTEI = Atome und Teilchen; MAE = Modelle adäquat einsetzen hinsichtlich Umgang mit mathematischen Grafiken; DisKont = Umgang mit Aspekten des Diskontinuums und Kontinuums; D = Diagnosemöglichkeiten auswählen können; H = Handlungsalternativen auswählen können; B = Bewertungen auswählen können; GF = Generalfaktor; NoS = Wissen über Nature of Science hinsichtlich Modellunterscheidungen treffen können; MWEMM = Modelle weiterentwickeln können und Modellmethode.

Der Ergebniszusammenfassung ist zu entnehmen, dass die Passung der empirischen Daten zu den theoretisch angenommenen Strukturen des Noticings und des Reasonings beim Biologietest über Fitindizes gezeigt werden kann. Dies ist im Bereich Chemie nicht der Fall bzw. nur unter Vorbehalten beim Noticing. Daher wurde dort nach einer passenden Struktur gesucht. Es ergab sich bei einer exploratorischen Faktorenanalyse ein sehr gut fittendes vierfaktorielles Modell (CFI/TLI: 0.998/0.996; RMSEA: 0.005). Die Benennung der Faktoren ist jedoch schwierig und scheint eine Mischung aus Noticing und Reasoning zu sein.

Es folgt eine Kurzzusammenfassung entlang der Hypothesen:

H1a-I) Die Passung der aus einem systematischen Literaturreview gewonnenen Struktur des Noticings (*Elemente zur Vermittlung von Modellkompetenz*) lässt sich im Biologietest belegen. Die Hypothese wird *nicht falsifiziert*.

H1a-II) Die Passung der aus einem systematischen Literaturreview gewonnenen Struktur des Noticings (*Elemente zur Vermittlung von Modellkompetenz*) lässt sich im Chemietest belegen. Die Hypothese wird *unter Vorbehalten nicht falsifiziert*.

H1b-I) Die Passung der Reasoningbestandteile (*aus Diagnosemöglichkeiten, Handlungsmöglichkeiten und Unterrichtsbewertungen auswählen können*) lässt sich im Biologietest belegen. Die Hypothese kann unter der Voraussetzung des Zusammenfalls von *Diagnose- und Handlungsmöglichkeiten auswählen können nicht falsifiziert* werden.

H1b-II) Die Passung der Reasoningbestandteile (*aus Diagnosemöglichkeiten, Handlungsmöglichkeiten und Unterrichtsbewertungen auswählen können*) lässt sich im Chemietest belegen. Die Hypothese muss unter der Voraussetzung des Zusammenfalls von allen drei Facetten *falsifiziert* werden.

H2a-I) Der Testscore aus dem Biologietest steigt querschnittlich in Abhängigkeit von der Anzahl der Fachsemester an. Die Hypothese kann *nicht falsifiziert* werden; es gibt sogar einen Zuwachs vom Ende des Studiums zum Referendariat hin.

H2a-II) Der Testscore aus dem Chemietest steigt querschnittlich in Abhängigkeit von der Anzahl der Fachsemester an. Die Hypothese kann *nicht falsifiziert* werden; ein signifikanter Zuwachs vom Ende des Studiums zum Referendariat hin bleibt jedoch aus.

H2b-I) Der Testscore im Biologietest steigt querschnittlich in Abhängigkeit von der Anzahl der geleisteten Unterrichtsstunden an. Die Hypothese muss *falsifiziert* werden.

H2b-II) Der Testscore im Chemietest steigt querschnittlich in Abhängigkeit von der Anzahl der geleisteten Unterrichtsstunden an. Die Hypothese muss *nicht falsifiziert* werden.

H2c-I) Die Antworten der Expertinnen und Experten der Biologiedidaktik bilden ein eindeutiges Antwortmuster als Referenzwert. Die Hypothese muss *nicht falsifiziert* werden.

H2c-II) Die Antworten der Expertinnen und Experten der Chemiedidaktik bilden ein eindeutiges Antwortmuster als Referenzwert. Die Hypothese muss *nicht falsifiziert* werden.

H3-I) Die Hauptaussagen der Ergebnisse aus dem Biologie- und Chemietest finden sich unabhängig vom Umgang mit fehlenden Werten wieder. Die Hypothese muss teilweise *falsifiziert*, teilweise *nicht falsifiziert* werden.

H3-II) Die Testschwierigkeit liegt unabhängig vom Umgang mit fehlenden Werten im mittleren Schwierigkeitsbereich. Die Hypothese kann *nicht falsifiziert* werden.

Erstmalig wurde auf Grundlage eines systematischen Literaturreviews ein theoriebasiertes Modell der *Elemente zur Vermittlung von Modellkompetenz* im Biologie- und Chemieunterricht entwickelt. Es gelang zum ersten Mal in einem Biologievignettentest, sowohl Noticing als auch Reasoning, also Professionelle Unterrichtswahrnehmung in ihrer Gesamtheit, zu modellieren. Die beim Chemievignettentest theoretisch angenommene Reasoning-Struktur findet sich nicht in den empirischen Daten. Die Noticingstruktur ergibt ein fittendes Modell; nur der Fitindex TLI ist knapp schlechter als gefordert. Die auf Grundlage einer exploratorischen Faktorenanalyse gefundene Struktur des Chemievignettentests fittet jedoch gut. Abschließend werden die für die Unterrichtspraxis besonders relevanten Ergebnisse genannt. So kann die Übersicht über die *Elemente zur Vermittlung von Modellkompetenz* zum einen zur Erstellung eines Curriculums, das Modellkompetenz explizit integriert, z. B. an Hochschulen genutzt werden. Zum anderen kann sie zur Operationalisierung weiterer Tests im Bereich Vermittlung von Modellkompetenz herangezogen werden. Außerdem können die Vignettentests für die Lehrkräfteaus- und -weiterbildung genutzt werden, um Probleme zu diagnostizieren und konkret an ihnen zu arbeiten. Des Weiteren können die Videovignetten die Arbeit in den Seminaren unterstützen, indem die gefilmten Unterrichtssequenzen den zukünftigen Lehrkräften die Möglichkeit bieten, den Modelleinsatz zu beobachten und zu reflektieren. Durch die Vereinheitlichung der Definition des Modellbegriffs und eine Übersicht über Definitionen zu Modelltypen sind einheitliche Begriffsnutzungen in Hochschule und Schule möglich.

8 Literaturverzeichnis

- Adobe. (2012) Adobe Illustrator CS6 [Computer software]: Adobe.
- Alonzo, A. C. & Kim, J. (2012). *Exploring Teachers' Pedagogical Content Knowledge Elicited with Video Clips Focused on Student Thinking. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Indianapolis, IN* (Annual meeting of the National Association for Research in Science Teaching) (S. 1-18).
- Angeli, C., Valanides, N. & Papageorgiou, M. (2011). An instructional design model for teaching pre-service elementary teachers to teach science with computer models. *Advanced Science Letters*, 4 (11-12), 3324-3331. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84856729345&partnerID=40&md5=893f2d9ec5f2c044b12617d6cde3509a>
- Artelt, C., Beinicke, A., Schlagmüller, M. & Schneider, W. (2009). Diagnose von Strategiewissen beim Textverstehen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 41 (2), 96-103.
- Aydin, S., Friedrichsen, P. M., Boz, Y. & Hanuscin, D. L. (2014). Examination of the topic-specific nature of pedagogical content knowledge in teaching electrochemical cells and nuclear reactions. *Chemistry Education Research and Practice*, 15 (4), 658-674. Zugriff am 26.05.2017. Verfügbar unter <http://pubs.rsc.org/en/content/articlehtml/2014/rp/c4rp00105b?page=search>
- Baer, M., Dörr, G., Fraefel, U., Kocher, M., Küster, O., Larcher, S. et al. (2007). Werden angehende Lehrpersonen durch das Studium kompetenter? Kompetenzaufbau und Standardreichung in der berufswissenschaftlichen Ausbildung an drei Pädagogischen Hochschulen in der Schweiz und in Deutschland. *Unterrichtswissenschaft*, 35 (1), 15-47. Verfügbar unter <http://www.pedocs.de/volltexte/2012/5485>; <http://nbn-resolving.de/urn:nbn:de:0111-opus-54857>
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Keyword: Teachers' professional competence]. *Zeitschrift für Erziehungswissenschaft*, 4, 469-520.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kom-*

- petenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29-53).
Münster [u.a.]: Waxmann.
- Baur, N. von & Florian, M. J. (2009). Stichprobenprobleme bei Online-Umfragen. In N. Jakob, H. Schönherr & T. Zerback (Hrsg.), *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung* (SpringerLink : Bücher, S. 109-128). Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH, Wiesbaden.
- Beerenwinkel, A. & Parchmann, I. (2008). Metadiskussionen über Modelle. Historische Aspekte als Impuls. *Praxis der Naturwissenschaften - Chemie in der Schule*, 57 (4), 13-16.
- Benedict, C. & Bolte, C. (2009). Wege in die Welt des Kleinen - Teilchenkonzepte von Grundschulkindern. In D. Höttecke (Hrsg.), *Chemie- und Physikdidaktik für die Lehramtsausbildung. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Schwäbisch Gmünd 2008* (Gesellschaft für Didaktik der Chemie und Physik#Bd, #29; Jahrestagung / Gesellschaft für Didaktik der Chemie und Physik. 2008, S. 458-460). Münster: Lit Verl.
- Bernholt, S., Parchmann, I. & Commons, M. L. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 219-245.
- Bielefeld, K. (2016). *Die beliebtesten Vornamen des Jahres 2000*. Zugriff am 12.07.2016.
Verfügbar unter <http://www.beliebte-vornamen.de/jahrgang/j2000>
- Bindernagel, J. A. & Eilks, I. (2009). Der "Roadmap-Ansatz zur Beschreibung und Entwicklung des Pedagogical Content Knowledge von Lehrkräften zum Teilchenkonzept. In D. Höttecke (Hrsg.), *Chemie- und Physikdidaktik für die Lehramtsausbildung. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Schwäbisch Gmünd 2008* (Gesellschaft für Didaktik der Chemie und Physik#Bd, #29; Jahrestagung / Gesellschaft für Didaktik der Chemie und Physik. 2008, S. 161-163). Münster: Lit Verl.
- Blömeke, S., Felbrich, A. & Müller, C. (2008). Theoretischer Rahmen und Untersuchungsdesign. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare - Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 15-48). Waxmann.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. Competence Viewed as a Continuum. *Zeitschrift für Psychologie*, 223 (1), 3-13.

- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare - Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung*: Waxmann.
- Blömeke, S., Kaiser, G., Schwarz, B., Seeber, S., Lehmann, R., Felbrich, A. et al. (2008). Fachbezogenes Wissen am Ende der Ausbildung. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare - Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 89-104). Waxmann.
- Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J., Döhrmann, M. et al. (2014). Von der Lehrerausbildung in den Beruf – Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 17 (3), 509-542.
- Blömeke, S., Lehmann, R., Seeber, S., Kaiser, G., Felbrich, A. & Müller, C. (2008). Niveau- und institutionenbezogene Modellierungen des fachbezogenen Wissens. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare - Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 105-134). Waxmann.
- Blömeke, S., Seeber, S., Lehmann, R., Kaiser, G., Schwarz, B. & Felbrich, A. (2008). Messung des fachbezogenen Wissens angehender Mathematiklehrkräfte [Measurement of the subject-specific knowledge of future mathematics teachers]. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare - Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 49-88). Waxmann.
- Blum, W. (2006). Modellierungsaufgaben im Mathematikunterricht - Herausforderung für Schüler und Lehrer. In A. Büchter, H. Humenberger, S. Hußmann, S. Prediger & H.-W. Henn (Hrsg.), *Realitätsnaher Mathematikunterricht - Vom Fach aus und für die Praxis. Festschrift für Hans-Wolfgang Henn zum 60. Geburtstag* (S. 8-23). Hildesheim [u.a.]: Franzbecker.

- Blum, W., Alsina, C., Biemengut, M. S., Bouleau, N., Confrey, J., Galbraith, P. et al. (2002). ICMI Study 14: Applications and Modelling in Mathematics Education – Discussion Document. *Zentralblatt für Didaktik der Mathematik (ZDM)*, 34 (5), 229-239.
- Bortz, J. & Döring, N. (Hrsg.). (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Mit 87 Tabellen* (Springer-Lehrbuch Bachelor, Master, 4., überarb. Aufl.). Berlin: Springer.
- Böttcher, D. M., Nitz, S. & Upmeyer zu Belzen, A. (2015, September). *Die professionelle Wahrnehmung von Experimentiersituationen im Biologieunterricht. Posterpräsentation auf der 20. Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO, Universität Hamburg.*
- Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, 71 (356), 791-799.
- Bromme, R. (1995). Was ist "pedagogical content knowledge"? Kritische Anmerkungen zu einem fruchtbaren Forschungsprogramm. *Zeitschrift für Pädagogik*, 33, 105-115.
- Bromme, R. (2004). Das implizite Wissen des Experten. In B. Koch-Priewe, F.-U. Kolbe & J. Wildt (Hrsg.), *Grundlagenforschung und mikrodidaktische Reformansätze zur Lehrerbildung* (S. 22-48). Bad Heilbrunn: Klinkhardt.
- Brovelli, D., Bölsterli, K., Rehm, M. & Wilhelm, M. (2013). Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht - ein Vignettest mit authentisch komplexen Unterrichtssituationen und offenem Antwortformat [Assessing Professional Competencies for Science Teaching: A Vignette Test using Authentically Complex Teaching Contexts and an Open-Ended Answer Format]. *Unterrichtswissenschaft*, 41, 306-329.
- Brovelli, D., Kauertz, A., Rehm, M. & Wilhelm, M. (2011). Professionelle Kompetenz und Berufsidentität in integrierten und disziplinären Lehramtsstudiengängen der Naturwissenschaften [Professional competence and professional identity in teacher training programs with integrated and separate science approaches]. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 57-87.
- Brunner, M., Kunter, M., Krauss, S., Klusmann, U., Baumert, J., Blum, W. et al. (2006). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht. Eine Zwischenbilanz des COACTIV-Projekts. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms : [BIQUA]*. Münster: Waxmann.

- Bühler, B. & Erb, R. (2010). Mikrokosmos und physikalisches Weltbild in den Vorstellungen der Schüler. In D. Höttecke (Hrsg.), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Dresden 2009 ; [36. Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik ; Gesellschaft für Didaktik der Chemie und Physik, Band 30]* (#30; Jahrestagung / Gesellschaft für Didaktik der Chemie und Physik. 2009, S. 83-85). Münster, Westf [u.a.]: LIT-Verl.
- Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42 (1), 116-131.
- Chalmers, P. (2014, 20. Februar). Package 'mirt'. *Multidimensional Item Response Theory*.
- Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48, 1-29.
- Christen, H. R. (1994). Das Orbital-Modell in der Schule? *Praxis der Naturwissenschaften. Chemie*, 43 (7), 29-31.
- Clarke, D. & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18 (8), 947-967.
- Cohen, R. & Yarden, A. (2009). Experienced junior-high-school teachers' PCK in light of a curriculum change: 'The cell is to be studied longitudinally'. *Research in Science Education*, 39 (1), 131-155. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-58549086117&partnerID=40&md5=fc0da4ce3852d452e8d3260623deae82>
- Coll, R. K. & Treagust, D. F. (2003). Investigation of secondary school, undergraduate, and graduate learners' mental models of ionic bonding. *Journal of Research in Science Teaching*, 40 (5), 464-486.
- Crawford, B. A. & Cullin, M. J. (2004). Supporting prospective teachers' conceptions of modelling in science. *International Journal of Science Education*, 26 (11), 1379-1401. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-10244266282&partnerID=40&md5=5fdb2254cd80342cd59462505a1fc558>
- Depaepe, F., Verschaffel, L. & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, 34, 12-25.

- Drechsler, M. & van Driel, J. H. (2008). Experienced teachers' pedagogical content knowledge of teaching acid-base chemistry. *Research in Science Education*, 38 (5), 611-631. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-54049122675&partnerID=40&md5=2a803b4d29598215a1e2362253628f8f>
- Duit, R. (2000). Konzeptwechsel und Lernen in den Naturwissenschaften in einem mehrperspektivistischen Ansatz. In R. Duit & C. v. Rhöneck (Hrsg.), *Ergebnisse fachdidaktischer und psychologischer Lehr-Lern-Forschung. Beiträge zu einem Workshop an der Pädagogischen Hochschule Ludwigsburg* (IPN, Bd. 169, S. 77-103). Kiel: Institut für die Pädagogik der Naturwissenschaften.
- Duit, R. H. & Treagust, D. F. (2012). Conceptual Change: Still a Powerful Framework for Improving the Practice of Science Instruction. In K. C. D. Tan & M. Kim (Hrsg.), *Issues and challenges in science education research. Moving forward* (S. 43-54). Dordrecht: Springer.
- Edelen, M. O. & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16 (S1), 5-18.
- Eilks, I., Witteck, T. & Pietzner, V. (2009a). Using Multimedia Learning Aids from the Internet for Teaching Chemistry. Not as Easy as it Seems? *Multiple Literacy and Science Education: ICTs in Formal and Informal Learning Environments*, 49-69. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84900179882&partnerID=40&md5=f156ec31a1799fd66c6d005524b63f85>
- Eilks, I., Witteck, T. & Pietzner, V. (2009b). Using multimedia learning aids from the internet for teaching chemistry: Not as easy as it seems? *Multiple Literacy and Science Education: ICTs in Formal and Informal Learning Environments*. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84900179882&partnerID=40&md5=f156ec31a1799fd66c6d005524b63f85>
- ETS (ETS, Hrsg.). (2015a). *Biology: Content Knowledge. The Praxis Study Companion. Measuring the Power of Learning*. Zugriff am 14.06.2016. Verfügbar unter <https://www.ets.org/praxis/prepare/materials/>
- ETS (Educational Testing Service, Hrsg.). (2015b). *Chemistry: Content Knowledge. The Praxis Study Companion. Listening. Learning. Leading*. Zugriff am 26.05.2017. Verfügbar unter <https://www.ets.org/praxis/prepare/materials/>

- Fend, H. (2001). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung* (Juventa-Paperback, 2., bereinigte Aufl.). Weinheim: Juventa-Verl.
- Fensham, P. J. (2004). *Defining an identity. The evolution of science education as a field of research* (Science & technology education library, Bd. 20). Dordrecht: Kluwer Academic.
- Freire, L. & Fernandez, C. (2014). Novice chemistry teachers and the development of PCK of oxidation-reduction: Influence of initial training [Professores novatos de química e o desenvolvimento do PCK de oxidorredução: Influências da formação inicial]. *Educacion Quimica*, 25 (3), 312-324. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84902777163&partnerID=40&md5=a07bc74e508339fae477acb8bd9a26bc>
- Gad, G. & Mittelsten Scheid, N. (2008). Seeberge. Modelle eines artenreichen ozeanischen Ökosystems. *Praxis der Naturwissenschaften - Biologie in der Schule*, 57 (1), 29-34.
- Garritz, A. (2013). Teaching the Philosophical Interpretations of Quantum Mechanics and Quantum Chemistry Through Controversies. *Science and Education*, 22 (7), 1787-1807. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84879955191&partnerID=40&md5=35a2d325bc89f7cf7b8ef31a3dd43a9c>
- Garson, G. D. (2015). *Missing Values Analysis & Data Imputation: Statistical Associates Publishing*. Verfügbar unter www.statisticalassociates.com
- Geiser, C. (2011). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung* (Lehrbuch, 2., durchges. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- George, D. & Mallery, P. (2003). *SPSS for Windows step by step. A simple guide and reference, 11.0 update* (4th ed.). Boston: Allyn and Bacon.
- Gerlitz, J.-Y. & Schupp, J. (2005). *Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. DIW Research Notes 2005* (Deutsches Institut für Wirtschaftsforschung (DIW), Hrsg.).
- Gess-Newsome, J. (1999). Pedagogical Content Knowledge: An Introduction And Orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge. The construct and its implications for science education* (Science & technology education library, vol. 6, pp. 3-17). Dordrecht [u.a.]: Kluwer Academic Publ.
- Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96, 606-633.

- Gramzow, Y., Riese, J. & Reinhold, P. (2013). Modellierung fachdidaktischen Wissens ange-
hender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 7-30.
- Graulich, N., Schreiner, P. R. & Hopf, H. (2010). Heuristic thinking makes a chemist smart.
In D. Höttecke (Hrsg.), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen
und Systematik. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Dres-
den 2009 ; [36. Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik ; Ge-
sellschaft für Didaktik der Chemie und Physik, Band 30] (#30; Jahrestagung / Gesellschaft
für Didaktik der Chemie und Physik. 2009, S. 449-451). Münster, Westf [u.a.]: LIT-Verl.*
- Gröger, M. (2010). Die historische Genese der "Minimumtonne" als Beispiel für die Entwick-
lung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik. In D. Höt-
tecke (Hrsg.), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Sys-
tematik. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Dresden 2009 ;
[36. Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik ; Gesellschaft für
Didaktik der Chemie und Physik, Band 30] (#30; Jahrestagung / Gesellschaft für Didaktik
der Chemie und Physik. 2009, S. 323-325). Münster, Westf [u.a.]: LIT-Verl.*
- Gruber, H. (2008). Lernen und Wissenserwerb. In W. Schneider & M. Hasselhorn (Hrsg.),
Handbuch der pädagogischen Psychologie (Bd. 10, S. 95-104). Göttingen: Hogrefe.
- Günther, S. L., Fleige, J., Upmeyer zu Belzen, A. & Krüger, D. (2017). Interventionsstudie mit
angehenden Lehrkräften zur Förderung von Modellkompetenz im Unterrichtsfach Biolo-
gie. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen
Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven* (S. 215-236).
Wiesbaden: Springer VS.
- Hammer, H. O. (1988). Symbole für Koordinationsverbindungen. *Naturwissenschaften im
Unterricht. Physik, Chemie* (35), 34-38.
- Harrison, A. & Jong, O. de. (2005). Using multiple analogies: Case study of a chemistry
teacher's preparations, presentations and reflections. *Research and the Quality of Science
Education*. Verfügbar unter [http://www.scopus.com/inward/record.url?eid=2-s2.0-
59149088073&partnerID=40&md5=e405f4a270a4b421d7ac83303b061b2e](http://www.scopus.com/inward/record.url?eid=2-s2.0-59149088073&partnerID=40&md5=e405f4a270a4b421d7ac83303b061b2e)
- Hattie, J. A. C. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to
achievement* (Reprinted). London [u.a.]: Routledge.
- Hattie, J. A. C. (2012). *Visible learning for teachers. Maximizing impact on learning*. Lon-
don: Routledge.

- Helsper, W. (2007). Eine Antwort auf Jürgen Baumerts und Mareike Kunters Kritik am strukturellen Professionsansatz. *Zeitschrift für Erziehungswissenschaft*, 10 (4), 567-579.
- Henkenborg, P., Krieger, A., Pinseler, J. & Behrens, R. (2008). Die Ebene des Unterrichts: Probleme der Unterrichtsgestaltung. In P. Henkenborg, A. Krieger, J. Pinseler & R. Behrens (Hrsg.), *Politische Bildung in Ostdeutschland. Demokratie-Lernen zwischen Anspruch und Wirklichkeit* (SpringerLink : Bücher, 1. Aufl., S. 39-101). Wiesbaden: VS, Verl. für Sozialwiss.
- Henze, I., van Driel, J. H. & Verloop, N. (2007). Science teachers' knowledge about teaching models and modelling in the context of a New Syllabus on Public understanding of science. *Research in Science Education*, 37 (2), 99-122. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-33947723272&partnerID=40&md5=5c0c92f513d29cd1517d97b3b8d9f7df>
- Henze, I., van Driel, J. H. & Verloop, N. (2008). Development of Experienced Science Teachers' Pedagogical Content Knowledge of Models of the Solar System and the Universe. *International Journal of Science Education*, 30 (10), 1321-1342.
- Hilgers, U., Sommerfeld, H. & Blume, R. (1994). Kronenether. Ein Carriermodell fuer den Ionentransport durch Biomembranen. *Der mathematische und naturwissenschaftliche Unterricht*, 47 (3), 179-184.
- Hill, H. C., Ball, D. L., Blunk, M., Goffney, I. M. & Rowan, B. (2007). Validating the ecological assumption: the relationship of measure scores to classroom teaching and student learning. *Measurement: Interdisciplinary Research and Perspectives*, 5, 107-118.
- Hopmann, S. & Riquarts, K. (1995). Didaktik and/or Curriculum: Basic Problems of Comparative Didaktik. In S. Hopmann & K. Riquarts (Eds.), *Didaktik and/or curriculum* (vol. 147, pp. 9-40). Kiel: Institut für die Pädagogik der Naturwissenschaften.
- Hu, L.-t. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.
- IBM Corporation. (2015) SPSS Statistics [Computer software].
- Ingham, A. M. & Gilbert, J. K. (1991). The use of analogue models by students of chemistry at higher education level. *International Journal of Science Education*, 13 (2), 193-202.

- Jahn, G., Prenzel, M., Stürmer, K. & Seidel, T. (2011). Varianten einer computergestützten Erhebung von Lehrkompetenzen: Untersuchungen zu Anwendungen des Tools Observer. *Unterrichtswissenschaft*, 39 (2), 136-153.
- Jahn, G., Stürmer, K., Seidel, T. & Prenzel, M. (2014). Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46 (4), 171-180.
- Jeffries, C. & Maeder, D. W. (2011). Comparing Vignette Instruction and Assessment Tasks to Classroom Observations and Reflections. *The Teacher Educator*, 46, 161-175.
- Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, 7 (2), 75-83.
- Jong, O. d., van Driel, J. H. & Verloop, N. (2005a). Preservice teachers' pedagogical content knowledge of using particle models in teaching chemistry. *Journal of Research in Science Teaching*, 42 (8), 947-964.
- Jong, O. d., van Driel, J. H. & Verloop, N. (2005b). Preservice teachers' pedagogical content knowledge of using particle models in teaching chemistry. *Journal of Research in Science Teaching*, 42 (8), 947-964. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-27844474260&partnerID=40&md5=d70f4279ebd97ad04e7580821782804b>
- Jong, O. de & van Driel, J. H. (2004). Exploring the development of student teachers' PCK of the multiple meanings of chemistry topics. *International Journal of Science and Mathematics Education*, 2 (4), 477-491. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-33644514961&partnerID=40&md5=dc573b883b3419dddc43f42f13287535>
- Justi, R. & van Driel, J. H. (2005a). A case study of the development of a beginning chemistry teacher's knowledge about models and modelling. *Research in Science Education*, 35 (2-3), 197-219. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-29444455175&partnerID=40&md5=ba9ffcadc85d41e2aa1f518d304e510>
- Justi, R. & van Driel, J. H. (2005b). Developing science teachers' knowledge on models and modelling. *Teacher Professional Development in Changing Conditions*. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84891992882&partnerID=40&md5=43a0b26919dae9ce20494fd38a4d53d2>

- Justi, R. & van Driel, J. H. (2005c). The development of science teachers' knowledge on models and modelling: Promoting, characterizing, and understanding the process. *International Journal of Science Education*, 27 (5), 549-573. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-18644383013&partnerID=40&md5=b5ed59b4a5af76a68549c555725a7d64>
- Jüttner, M. (2012). *Entwicklung, Evaluation und Validierung eines Fachwissenstests und eines fachdidaktischen Wissenstests für die Erfassung des Professionswissens von Biologielehrkräften*.
- Jüttner, M. & Neuhaus, B. J. (2013). Das Professionswissen von Biologielehrkräften – Ein Vergleich zwischen Biologielehrkräften, Biologen und Pädagogen. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 31-49.
- Kaiser, G. (2016, März). *Videobasierte Erfassung professioneller Kompetenzen von Lehrkräften bei der Förderung von Kreativität und mathematischer Begabung – Detailergebnisse aus TEDS-FU*, Heidelberg.
- Kattmann, U. (2015). *Schüler besser verstehen. Alltagsvorstellungen im Biologieunterricht*. Hallbergmoos: Aulis.
- Kersting, M., Althoff, K. & Jäger, A. O. (2008). *WIT-2. Wilde-Intelligenz-Test 2*. Göttingen: Hogrefe.
- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R. & Stigler, J. W. (2012). Measuring Usable Knowledge: Teachers' Analyses of Mathematics Classroom Videos Predict Teaching Quality and Student Learning. *American Educational Research Journal*, 49, 568-589.
- Keune, H. & Daemngen, U. (1988). Zu den Formeldarstellungen des Glucose-Moleküls. *Der mathematische und naturwissenschaftliche Unterricht*, 41 (1), 8-12.
- Khourey-Bowers, C. & Fenk, C. (2009). Influence of constructivist professional development on chemistry content knowledge and scientific model development. *Journal of Science Teacher Education*, 20 (5), 437-457. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-76149116506&partnerID=40&md5=81d2427c4a4fce6338a62eae21396410>
- Kirschner, S. (2013). *Modellierung und Analyse des Professionswissens von Physiklehrkräften* (Studien zum Physik- und Chemielernen, Bd. 161).

- Kleinknecht, M. (2014, Februar). *Unterrichtsanalysekompetenz fördern*. Kolloquium, Ludwigsburg.
- Kleinknecht, M. & Schneider, J. (2013). What do teachers think and feel when analyzing videos of themselves and other teachers teaching? *Teaching and Teacher Education*, 33, 13-23.
- Klinger, W. (2001). Bedeutung und Möglichkeiten der Veranschaulichung im Bereich der Physik. In J. Forster & U. Krebs (Hrsg.), *Das "praktische Lernen" und das Problem der Wissenskumulation. Von der Notwendigkeit der Entwicklung neuer Niveaus der Anschaulichkeit*. (Schriftenreihe zum Bayerischen Schulmuseum Ichenhausen, Zweigmuseum des Bayerischen Nationalmuseums, und zum Schulmuseum Nürnberg, Bd. 20, S. 167-204). Bad Heilbrunn: Klinkhardt.
- KMK Standards (Kultusministerkonferenz, Hrsg.). (2005a). *Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. Zugriff am 14.10.2013. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Biologie.pdf
- KMK Standards (Kultusministerkonferenz, Hrsg.). (2005b). *Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. Zugriff am 14.10.2013. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Chemie.pdf
- König, A. & Reiners, C. S. (2004). Computergestützte Lehr- und Lernmaterialien zur chemischen Bindung. In A. Pitton (Hrsg.), *Chemie- und physikdidaktische Forschung und naturwissenschaftliche Bildung* (Jahrestagung / Gesellschaft für Didaktik der Chemie und Physik#Bd, #2003; Gesellschaft für Didaktik der Chemie und Physik. 24, S. 334-336). Münster: Lit.
- König, J. & Blömeke, S. (2010). *Pädagogisches Unterrichtswissen (PUW). Dokumentation der Kurzfassung des TEDS-M Testinstruments zur Kompetenzmessung in der ersten Phase der Lehrerausbildung*: Humboldt-Universität zu Berlin.
- Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M. et al. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neu-

- brand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 135-161). Münster [u.a.]: Waxmann.
- Krell, M., Reinisch, B. & Krüger, D. (2014). Analyzing Students' Understanding of Models and Modeling Referring to the Disciplines Biology, Chemistry, and Physics. *Research in Science Education*.
- Kromrey, J. D. & Renfrow, D. D. (1991). Using Multiple Choice Examination Items To Measure Teachers' Content-Specific Pedagogical Knowledge. In Eastern Educational Research Association (Hrsg.), *Annual Meeting of the Eastern Educational Research Association (Boston, Massachusetts)* (S. 1-20).
- Kube, J. (2009). Vornamensforschung: Fragebogenuntersuchung bei Lehrerinnen und Lehrern, ob Vorurteile bezüglich spezifischer Vornamen von Grundschulern und davon abgeleitete erwartete spezifische Persönlichkeitsmerkmale vorliegen. *Master Arbeit, Universität Oldenburg* <http://www.kinderforschung.uni-oldenburg.de/36968.html>.
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts* (StandardWissen Lehramt, 1., neue Ausg). Paderborn: UTB; Schöningh Paderborn.
- Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85-113). Münster [u.a.]: Waxmann.
- Leiner, D. & Leiner, S. (SoSci Survey GmbH, Hrsg.). (2015). *SoSciSurvey.de. oFB - der Onlinefragebogen*, SoSci Survey GmbH. Verfügbar unter <https://www.soscisurvey.de/>
- Leisner, A. & Mikelskis, H. F. (2004). Erwerb metakonzeptueller Kompetenz durch ein systematisches Lernen über Modelle. In A. Pitton (Hrsg.), *Chemie- und physikdidaktische Forschung und naturwissenschaftliche Bildung* (Jahrestagung / Gesellschaft für Didaktik der Chemie und Physik#Bd, 2003; Gesellschaft für Didaktik der Chemie und Physik. 24, S. 120-122). Münster: Lit.
- Leisner-Bodenthin, A. (2006). Zur Entwicklung von Modellkompetenz im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 91-109.
- Lethmate, J. & Arning, H. (2003). Experimentelles Arbeiten im Geographieunterricht. *Geographie und Schule*, 25 (145), 35-39.

- Liang, J.-C., Chou, C.-C. & Chiu, M.-H. (2011). Student test performances on behavior of gas particles and mismatch of teacher predictions. *Chemistry Education Research and Practice*, 12 (2), 238-250. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-80054094650&partnerID=40&md5=6eefa43b56bc5f6b40ee5f4ba737099f>
- Lin, J.-W. & Chiu, M.-H. (2010). The mismatch between students' mental models of acids/bases and their sources and their teacher's anticipations thereof. *International Journal of Science Education*, 32 (12), 1617-1646. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-77954495760&partnerID=40&md5=bc93071c37b11dc6744e6e1980770dd9>
- Litsche, G. (1985). Arbeit mit gedanklichen Modellen im Unterricht. *Biologie in der Schule*, 34 (7-8), 262-266.
- Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83 (404), 1198-1202.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58 (2), 103-117.
- Markert, B. (1988). Ein einfaches Synapsenmodell zur Simulation erregender Transmitterschütterungen und zur Erklärung der Wirkungsweise einiger Nervengifte fuer den Biologieunterricht in der Sekundarstufe II. *Der mathematische und naturwissenschaftliche Unterricht*, 41 (3), 168-171.
- Marx, A. & Rinkens, H.-D. (2008). Anforderungsmerkmale der MT21-Items und ihre Weiterentwicklung aus mathematikdidaktischer Sicht. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare - Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 425-452). Waxmann.
- McClelland, D. C. (1973). Testing for competence rather than for "intelligence". *American Psychologist*, 28, 1-14.
- Meisert, A. (2008). Vom Modellwissen zum Modellverständnis - Elemente einer umfassenden Modellkompetenz und deren Fundierung durch lernseitige Kriterien zur Klassifikation von Modellen. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 243-261. Zugriff am 18.10.2013. Verfügbar unter http://www.ipn.uni-kiel.de/zfdn/pdf/14_013_Meisert.pdf
- Meisert, A. (2014). Modellverständnis fördern. *Unterricht Biologie* (397/398), 76-77.

- Meschede, N. (2014). *Professionelle Wahrnehmung der inhaltlichen Strukturierung im naturwissenschaftlichen Grundschulunterricht. Theoretische Beschreibung und empirische Erfassung* (Bd. 163). Berlin: Logos-Verl.
- Meschede, N., Steffensky, M., Wolters, M. & Möller, K. (2015). Professionelle Wahrnehmung der Lernunterstützung im naturwissenschaftlichen Grundschulunterricht. *Unterrichtswissenschaft*, 43 (4), 317-335.
- Messick, S. (1989). Validity. In R. L. Linn (Hrsg.), *Educational measurement* (The American Council on Education, 3. Aufl., S. 13-103). New York: Collier Macmillan Publishers.
- Meyer, H. (1990). Modelle. *Unterricht Biologie*, 14 (160), 4-10.
- Mikelskis-Seifert, S. (2004). Über Modelle lernen - Empirische Erforschung einer theoriegeleiteten Konzeption und Konsequenzen für die Unterrichtspraxis. In A. Pitton (Hrsg.), *Chemie- und physikdidaktische Forschung und naturwissenschaftliche Bildung* (Jahrestagung / Gesellschaft für Didaktik der Chemie und Physik#Bd, #2003; Gesellschaft für Didaktik der Chemie und Physik. 24, S. 14-29). Münster: Lit.
- Mikelskis-Seifert, S. & Leisner, A. (2004). Lernen über Modelle in einer interdisziplinären Projektwoche der 8. Klassen. In A. Pitton (Hrsg.), *Chemie- und physikdidaktische Forschung und naturwissenschaftliche Bildung* (Jahrestagung / Gesellschaft für Didaktik der Chemie und Physik#Bd, #2003; Gesellschaft für Didaktik der Chemie und Physik. 24, S. 123-125). Münster: Lit.
- Mikelskis-Seifert, S., Thiele, M. & Wünsch, T. (2005). Modellieren - Schlüsselfähigkeit für physikalische Forschungs- und Lernprozesse. *PhyDid A, Physik und Didaktik in Schule und Hochschule*, 4 (1), 30-46. Verfügbar unter <http://www.phydid.de/index.php/phydid/article/view/29>
- Ministerium für Kultus, Jugend und Sport (Hrsg.). (2012). *Bildungsplan 2012. Werkrealschule*. Zugriff am 18.10.2013. Verfügbar unter http://www.bildung-staerkt-men-schen.de/service/downloads/Bildungsplaene/Werkrealschule/Bildungsplan2012_WRS_Internet.pdf
- Muthén, L. K. & Muthén, B. O. (2012). *Mplus User's Guide* (Seventh Edition). Los Angeles, CA: Muthén & Muthén.

- Nehring, A., Nowak, K. H., Upmeyer zu Belzen, A. & Tiemann, R. (2015). Predicting Students' Skills in the Context of Scientific Inquiry with Cognitive, Motivational, and Socio-demographic Variables. *International Journal of Science Education*, 37 (9), 1343-1363.
- Nelson, M. M. & Davis, E. A. (2012). Preservice Elementary Teachers' Evaluations of Elementary Students' Scientific Models: An aspect of pedagogical content knowledge for scientific modeling. *International Journal of Science Education*, 34 (12), 1931-1959. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84864657765&partnerID=40&md5=7222ddc80657c4ae842930e9350cab4>
- Niss, M. (2012). Models and Modelling in Mathematics Education. *European Mathematical Society Newsletter*, 49-52.
- Nocke, H. (2001). Das Gegenstromprinzip. Teil 1: Ein Spielmodell für aktive Gegenstromsysteme. *Praxis der Naturwissenschaften - Biologie in der Schule*, 50 (2), 39-42.
- Oevermann, U. (1996). Theoretische Skizze einer revidierten Theorie professionalisierten Handelns. In A. Combe & W. Helsper (Hrsg.), *Pädagogische Professionalität. Untersuchungen zum Typus pädagogischen Handelns* (Suhrkamp Taschenbuch Wissenschaft, Bd. 1230, 1. Aufl., S. 70-182). Frankfurt am Main: Suhrkamp.
- Oh, P. S. & Kim, K. S. (2013). Pedagogical Transformations of Science Content Knowledge in Korean Elementary Classrooms. *International Journal of Science Education*, 35 (9), 1590-1624. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84879696589&partnerID=40&md5=32d5ede2dd55750fadedd13d1fe671ab>
- Okanlawon, A. E. (2010). Constructing a framework for teaching reaction stoichiometry using pedagogical content knowledge. *Chemistry*, 19 (2), E27-E44. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-82055195328&partnerID=40&md5=f51b4d67f02bf371434eb175c8085829>
- Oser, F., Heinzer, S. & Salzmann, P. (2010). Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. *Unterrichtswissenschaft*, 38, 5-27.
- Padilla, K., Ponce-de-León, A. M., Rembado, F. M. & Garritz, A. (2008). Undergraduate professors' pedagogical content knowledge: The case of 'amount of substance'. *International Journal of Science Education*, 30 (10), 1389-1404. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-48449092563&partnerID=40&md5=911514d06440713102bb6838d848eac>

- Papageorgiou, G., Stamovlasis, D. & Johnson, P. M. (2010). Primary teachers' particle ideas and explanations of physical phenomena: Effect of an in-service training course. *International Journal of Science Education*, 32 (5), 629-652. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-77951144254&partnerID=40&md5=a66f67c767c4c19c0b26f5b098e76ed1>
- Parchmann, I. (2012). Schülervorstellungen - Lernbarrieren oder Lernchancen? *MNU Der mathematische und naturwissenschaftliche Unterricht*, 26-29.
- Park, E. J. & Light, G. (2009). Identifying Atomic Structure as a Threshold Concept. Student mental models and troublesomeness. *International Journal of Science Education*, 31 (2), 233-258.
- Peters, W. (1998). *Zur Theorie der Modellierung von Natur und Umwelt*. : Technische Universität Berlin.
- Pohlmann, B. & Möller, J. (2010). Fragebogen zur Erfassung der Motivation für die Wahl des Lehramtsstudiums (FEMOLA). *Zeitschrift für Pädagogische Psychologie*, 24 (1), 73-84.
- Posner, G. J., Strike, K. A., Hewson, P. W. & Gertzog, W. A. (1982). Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change. *Science Education*, 66, 211-227.
- Rehm, M. (2007). Crossing the bridge from the macro to the micro world. *NVOX - Tijdschrift voor natuurwetenschap op school*, 32 (6), 291-293.
- Rehm, M. & Bölsterli, K. (2014). Entwicklung von Unterrichtsvignetten. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 213-225). Berlin, Heidelberg: Springer.
- Retelsdorf, J., Bauer, J., Gebauer, S. K., Kauper, T. & Möller, J. (2014). Erfassung berufsbezogener Selbstkonzepte von angehenden Lehrkräften (ERBSE-L). *Diagnostica*, 60 (2), 98-110.
- Riehs, N. F. (2013). *Epistemologische Einstellungen von angehenden Chemie- und Physiklehrern. Wissenschaftlicher Realismus und das Lehren von Modellen*. Zugriff am 18.12.2013.
- Riese, J. & Reinhold, P. (2010). Empirische Erkenntnisse zur Struktur professioneller Handlungskompetenz von angehenden Physiklehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 167-187.

- Rindermann, H. & Neubauer, A. C. (2000). Informationsverarbeitungsgeschwindigkeit und Schulerfolg: Weisen basale Maße der Intelligenz prädiktive Validität auf. *Diagnostica*, 46 (1), 8-17.
- Roberts, P. D., Stewart, G. B. & Pullin, A. S. (2006). Are review articles a reliable source of evidence to support conservation and environmental management? A comparison with medicine. *Biological Conservation*, 132 (4), 409-423.
- Romine, W. L. & Walter, E. M. (2014). Assessing the efficacy of the measure of understanding of macroevolution as a valid tool for undergraduate non-science majors. *International Journal of Science Education*, 36 (17), 2872-2891. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84926244398&partnerID=40&md5=b497e502e49bc92623f8511f346f5637>
- Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion* (Aus dem Programm Huber: Psychologie-Lehrbuch, 1. Aufl.). Bern [u.a.]: Huber.
- Rost, J. (2004). *Lehrbuch Testtheorie, Testkonstruktion*. Bern: Hans Huber.
- Rueckl, E. & Ebinghaus, H. (1989). Genese der Atommodelle mit dem Feldenergiekonzept. *Praxis der Naturwissenschaften. Physik*, 38 (4), 21-22.
- Saborowski, J. (2000). *Computervisualisierung und Modelldenken. Konzeptionelle Grundlagen und fachdidaktische Konsequenzen für den Chemieunterricht*. Köln: Saborowski.
- Santagata, R. & Yeh, C. (2015). The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM*, 153-165.
- Schmelzing, S., van Driel, J. H., Jüttner, M., Brandenbusch, S., Sandmann, A. & Neuhaus, B. J. (2013a). DEVELOPMENT, EVALUATION, AND VALIDATION OF A PAPER- AND-PENCIL TEST FOR MEASURING TWO COMPONENTS OF BIOLOGY TEACHERS' PEDAGOGICAL CONTENT KNOWLEDGE CONCERNING THE "CARDIOVASCULAR SYSTEM. *International Journal of Science and Mathematics Education*, 11 (6), 1369-1390. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84890427720&partnerID=40&md5=04229100e6aa8596501020162c4ec98e>
- Schmelzing, S., van Driel, J. H., Jüttner, M., Brandenbusch, S., Sandmann, A. & Neuhaus, B. J. (2013b). DEVELOPMENT, EVALUATION, AND VALIDATION OF A PAPER- AND-PENCIL TEST FOR MEASURING TWO COMPONENTS OF BIOLOGY TEACHERS' PEDAGOGICAL CONTENT KNOWLEDGE CONCERNING THE

- “CARDIOVASCULAR SYSTEM”. *International Journal of Science and Mathematics Education, 11* (6), 1369-1390. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84890427720&partnerID=40&md5=04229100e6aa8596501020162c4ec98e>
- Schmelzing, S., Wüsten, S., Sandmann, A. & Neuhaus, B. J. (2010). Fachdidaktisches Wissen und Reflektieren im Querschnitt der Biologielehrerbildung. *Zeitschrift für Didaktik der Naturwissenschaften, 16*, 189-207.
- Schmidt, W. H., Tatto, M. T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L. et al. (2008). How Does the German Sample in Comparison to the Samples from Bulgaria, Mexico, South Korea, Taiwan and the US? In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare - Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 481-498). Waxmann.
- Schneider, H. (1981). Quantifizierender Einsatz von Modellen im Biologieunterricht - aufgezeigt am Beispiel der menschlichen Wirbelsäule. *Naturwissenschaften im Unterricht. Biologie, 29* (8), 259-263.
- Schrader, N. (2015). *Schülervorstellungen im Biologie-Unterricht. Das Modell der Didaktischen Rekonstruktion (MDR)*. München: GRIN Verlag.
- Seidel, T., Blomberg, G. & Stürmer, K. (2010). "Observer" - Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. In E. Klieme, D. Leutner & M. Kenk (Hrsg.) *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*, 296-306 [Themenheft]. Beltz. Verfügbar unter <http://www.pedocs.de/volltexte/2010/3438>; <http://nbn-resolving.de/urn:nbn:de:0111-opus-34384>
- Seidel, T., Prenzel, M., Rimmel, R., Dalehefte, I. M., Herweg, C., Kobarg, M. et al. (2006). Blicke auf den Physikunterricht. Ergebnisse der IPN Videostudie. *Zeitschrift für Pädagogik, 52* (6), 799-821.
- Seidel, T. & Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal, 51* (4), 739-771.
- Sherin, M. G. (2007). The Development of Teachers' Professional Vision in Video Clubs. In R. Goldman, P. Pea, B. Barron & S. J. Derry (Hrsg.), *Video research in the learning sciences* (S. 383-395). Mahwah, N.J.: Lawrence Erlbaum Associates, Inc.

- Sherin, M. G. & van Es, E. A. (2009). Effects of Video Club Participation on Teachers' Professional Vision. *Journal of Teacher Education*, 60, 20-37.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4-14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, 1-21.
- Shulman, L. S. (1991). Von einer Sache etwas verstehen: Wissensentwicklung bei Lehrern. In E. Terhart (Hrsg.), *Unterrichten als Beruf. Neuere amerikanische und englische Arbeiten zur Berufskultur und Berufsbiographie von Lehrern und Lehrerinnen* (S. 145-160). Köln: Böhlau.
- Šorgo, A., Usak, M., Kubiato, M., Fančovičova, J., Prokop, P., Puhek, M. et al. (2014). A cross-cultural study on freshmen's knowledge of genetics, evolution, and the nature of science. *Journal of Baltic Science Education*, 13 (1), 6-18. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84900437073&partnerID=40&md5=7835171c506f6145bab9bdf4a0546988>
- Sprinthall, N. A., Reiman, A. J. & Thies-Sprinthall, L. (1996). Teacher professional development. In J. Sikula, T. J. Buttery & E. Guyton (Hrsg.), *Handbook of research on teacher education*. New York: Macmillan.
- Stachowiak, H. (1980). Einleitung: Der Weg zum Systemischen Neopragmatismus und das Konzept der Allgemeinen Modelltheorie. In H. Stachowiak (Hrsg.), *Modelle und Modelldenken im Unterricht. Anwendungen der allgemeinen Modelltheorie auf die Unterrichtspraxis* (Forschen und Lernen, Bd. 4, 4. Aufl., Bd. 4, S. 9-49). Bad Heilbrunn/Obb.: Klinkhardt. Zugriff am 25.03.2013.
- Statistische Ämter des Bundes und der Länder. (2011). *Bevölkerung nach Migrationsstatus regional 2011*.
- Stecher, B., Le, V.-N., Hamilton, L., Ryan, G., Robyn, A. & Lockwood, J. R. (2006). Using Structured Classroom Vignettes to Measure Instructional Practices in Mathematics. *Educational Evaluation and Policy Analysis*, 28 (2), 101-130.
- Steffensky, M., Gold, B., Holdynski, M. & Möller, K. (2015). Professional Vision of Classroom Management and Learning Support in Science Classrooms—Does Professional Vision Differ Across General and Content-Specific Classroom Interactions? *International Journal of Science and Mathematics Education*, 13 (2), 351-368.

- Steinbeck, R. & Markert, B. (1989). Ansätze zur Arbeit mit dem Regelkreismodell. *Praxis der Naturwissenschaften. Biologie*, 38 (4), 39-41.
- Streller, S. & Bolte, C. (2008). Tornados im globalen Klimageschehen. Eine fachübergreifende Sequenz für den Chemieanfangsunterricht. *Praxis der Naturwissenschaften - Chemie in der Schule*, 57 (2), 11-16.
- Strübe, M., Tröger, H., Tepner, O. & Sumfleth, E. (2014). Development of a Pedagogical Content Knowledge test of chemistry language and models [Desarrollo de una prueba de Conocimiento Pedagógico del Contenido sobre modelos y el lenguaje de la química]. *Educacion Quimica*, 25 (3), 380-390. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84902803996&partnerID=40&md5=07e6f0d6e0c631edf796391a5e18fd95>
- Stürmer, K. & Seidel, T. (2015). Assessing Professional Vision in Teacher Candidates. *Zeitschrift für Psychologie*, 223 (1), 54-63.
- Sunder, C., Todorova, M. & Möller, K. (2015). Kann die professionelle Unterrichtswahrnehmung von Sachunterrichtsstudierenden trainiert werden? – Konzeption und Erprobung einer Intervention mit Videos aus dem naturwissenschaftlichen Grundschulunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*.
- Syring, M., Bohl, T., Kleinknecht, M., Kuntze, S., Rehm, M. & Schneider, J. (2015). Videos oder Texte in der Lehrerbildung? Effekte unterschiedlicher Medien auf die kognitive Belastung und die motivational-emotionalen Prozesse beim Lernen mit Fällen. *Zeitschrift für Erziehungswissenschaft*, 18 (4), 667-685.
- Tepner, O. & Dollny, S. (2014). Entwicklung eines Testverfahrens zur Analyse fachdidaktischen Wissens. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 311-323). Berlin, Heidelberg: Springer.
- Thillmann, H. (2008). *Selbstreguliertes Lernen durch Experimentieren: Von der Erfassung zur Förderung*. Essen. Verfügbar unter <http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=17311>
- Touché, W. (1989). Atommodelle im Chemie- und Physikunterricht der gymnasialen Mittelstufe. Versuch e. Koordination. *Praxis der Naturwissenschaften. Physik*, 38 (4), 2-6.
- Trauschke, M. (2008). Krieg im menschlichen Körper - Überfälle böser Viren. *MNU Der mathematische und naturwissenschaftliche Unterricht*, 61 (8), 493-499.

- Unal, S., Sadoglu, G. P. & Durukan, U. G. (2014). Teacher educators' views of "model" concept and their mental models. *Journal of Baltic Science Education*, 13 (5), 674-694. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84923266046&partnerID=40&md5=f24c16ec1b00391b6ee6d0e25bc0b436>
- Upmeier zu Belzen, A. & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41-57. Zugriff am 18.10.2013. Verfügbar unter http://www.ipn.uni-kiel.de/zfdn/pdf/16_Upmeier.pdf
- Usak, M., Ozden, M. & Eilks, I. (2011). A case study of beginning science teachers' subject matter (SMK) and pedagogical content knowledge (PCK) of teaching chemical reaction in Turkey. *European Journal of Teacher Education*, 34 (4), 407-429. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-80053154878&partnerID=40&md5=996e83c076d7775c15112ca990ba2f31>
- Valanides, N. & Angeli, C. (2006). Preparing preservice elementary teachers to teach science through computer models. *Contemporary Issues in Technology and Teacher Education*, 6 (1), 87-98.
- Van Dijk, E. M. (2009). Teachers' views on understanding evolutionary theory: A PCK-study in the framework of the ERTE-model. *Teaching and Teacher Education*, 25 (2), 259-267. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-59149084075&partnerID=40&md5=45c23871bd4480123fa9114ff9d1e99f>
- Van Dijk, E. M. & Kattmann, U. (2007). A research model for the study of science teachers' PCK and improving teacher education. *Teaching and Teacher Education*, 23 (6), 885-897. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-34447117715&partnerID=40&md5=cfbe308fa6459259405d75d14bf2ae48>
- Van Driel, J. H. & Jong, O. d. (2015). Empowering Chemistry Teachers' Learning: Practices and New Challenges. In J. García-Martínez & E. Serrano-Torregrosa (Hrsg.), *Chemistry education. Best practices, opportunities and trends* (S. 2-22).
- Van Driel, J. H., Jong, O. d. & Verloop, N. (2002). The development of preservice chemistry teachers' pedagogical content knowledge. *Science Education*, 86 (4), 572-590. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-0141992101&partnerID=40&md5=396135c811744f411158716e177e67d3>
- Van Driel, J. H., Jong, O. de & Verloop, N. (2002). The Development of Preservice Chemistry Teachers' Pedagogical Content Knowledge. *Science Education*, 86 (4), 572-590. Ver-

- füßbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-0141992101&partnerID=40&md5=396135c811744f411158716e177e67d3>
- Van Driel, J. H. & Verloop, N. (1999). Teachers' knowledge of models and modelling in science. *International Journal of Science Education*, 21, 1141-1153.
- Van Driel, J. H. & Verloop, N. (2002). Experienced teachers' knowledge of teaching and learning of models and modelling in science education. *International Journal of Science Education*, 24, 1255-1272.
- Van Es, E. A. & Sherin, M. G. (2002). Learning to Notice: Scaffolding New Teachers' Interpretations of Classroom Interactions. *Journal of Technology and Teacher Education*, 10 (4), 571-596.
- Voss, T. & Kunter, M. (2011). Pädagogisch-psychologisches Wissen von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 193-214). Münster [u.a.]: Waxmann.
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103 (4), 952-969.
- Wang, Z., Chi, S., Hu, K. & Chen, W. (2014). Chemistry Teachers' Knowledge and Application of Models. *Journal of Science Education and Technology*, 23 (2), 211-226. Verfügbar unter <http://www.scopus.com/inward/record.url?eid=2-s2.0-84897662146&partnerID=40&md5=b67515c556d558223a0b94585c2d3ebb>
- Weinert, F. E. (2001). Concept of Competence: A Conceptual Clarification. In D. S. Rychen & L. H. Sagalnik (Hrsg.), *Definition and selection of competencies--theoretical and conceptual foundations* (S. 45-65). Kirkland, WA: Hogrefe & Huber.
- Wikipedia (Hrsg.). (2016, 13. Februar). *Vignette*. Zugriff am 16.03.2016. Verfügbar unter <https://de.wikipedia.org/w/index.php?oldid=148505502>
- Wilhelm, O. & Nickolaus, R. (2013). Was grenzt das Kompetenzkonzept von etablierten Kategorien wie Fähigkeit, Fertigkeit oder Intelligenz ab? *Zeitschrift für Erziehungswissenschaft*, 16 (S1), 23-26.
- Winnenburg, W. (2000). Der Sternenhimmel - Vom Phänomen zum Modell. In R. Brechel (Hrsg.), *Vorträge auf der Tagung für Didaktik der Physik, Chemie. In München, September*

1999 (Zur Didaktik der Physik und Chemie, Bd. 20, Tagung 1999, S. 324-326). Alsbach/Bergstrasse: Leuchtturm-Verlag.

Witner, S. & Tepner, O. (2011a). Entwicklung geschlossener Testaufgaben zur Erhebung des fachdidaktischen Wissens von Chemielehrkräften. *chimica et ceterae artes rerum naturae didacticae*, 37, 113-137.

Witner, S. & Tepner, O. (2011b). Entwicklung geschlossener Testaufgaben zur Erhebung des fachdidaktischen Wissens von Chemielehrkräften. *chimica et ceterae artes rerum naturae didacticae*, 37, 113-137.

Witner, S. & Tepner, O. (2011c). ProwiN-Test zum Fachwissen von Chemielehrkräften. In A. Borowski, H. E. Fischer, M. Jüttner, S. Kirschner, B. J. Neuhaus, E. Sumfleth et al. (Hrsg.), *ProwiN-Testinstrumente..* Essen: Universität Duisburg-Essen.

9 Anhang

9.1 Eidesstattliche Versicherung

Eidesstattliche Versicherung gem. § 7 Abs. 2 (e) der Promotionsordnung der Pädagogischen Hochschule Heidelberg

1. Bei der eingereichten Dissertation zu dem Thema

Vermittlung von Modellkompetenz in den Unterrichtsfächern Biologie und Chemie. Modellierung, Validierung und Messung Professioneller Unterrichtswahrnehmung zukünftiger Lehrkräfte mit Hilfe eines Vignettentests

handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Heidelberg, den _____

(Unterschrift Benjamin J. Tempel)

9.2 Tabellen zum systematischen Literaturreview

In den folgenden, die Ergebnisse des systematischen Literaturreview zusammenfassenden, vier Tabellen wird bei klein beginnenden Zitaten aus Platzgründen auf die Auslassungszeichen (...) verzichtet.

9.2.1 Überblick über Art des Artikels, Modelldefinition, Thematisierung von Modellen und Herkunftsland

		Quelle	Art des Artikels (z. B. normativ, theoretisch, empirisch: qualitativ, quantitativ)	Explizite Definition des Modellbegriffs	Elemente von Modellen im Unterricht (Metabetrachtung von Modellen vs. Betrachtung eines spezifischen Modells)	Das Land, aus dem die Studie kommt bzw. in dem die Daten erhoben wurden
1	1	Beerenwinkel und Parchmann (2008)	Theoretisch, Unterrichtspraxis	<ul style="list-style-type: none"> - Ein Modell löst „eine Interpretation eines empirischen Phänomens aus, und zwar auf eine Weise, die den (intellektuellen) Zugang zu diesem Phänomen erleichtert. Interpretationen kommen zustande z. B. durch Idealisierung, Vereinfachung oder Analogisieren.“ (S. 13), Boiler-Jones 2000, 1 f. - Aufzählung von Eigenschaften (van Driel & Verloop, 1999, S. 1142), die allen Modellen gemein sind. - „Ein Modell steht immer in Beziehung zu einem Original, welches durch das Modell repräsentiert wird.“ (S. 13) - „Ein Modell ist ein Hilfsmittel des Wissenschaftlers, um Informationen über Originale zu erhalten, die nicht direkt beobachtbar oder messbar sind.“ (S. 13) - „Ein Modell kann nicht direkt mit dem Original, das es repräsentiert, interagieren.“ (S. 13) - „Ein Modell weist bestimmte Analogien zu dem Original auf.“ (S. 13) - „Ein Modell unterscheidet sich immer in bestimmten Aspekten von dem Original.“ (S. 13) - „Bei dem Entwurf eines Modells muss immer ein Kompromiss zwischen den Analogien und Unterschieden 	Historische Aspekte als Impulsgeber für Metadiskussionen über chemische Modelle	Deutschland

				gefunden werden.“ (S. 13) - „Die Entwicklung eines Modells ist ein iterativer Prozess.“ (S. 13)		
2	1	Benedict und Bolte (2009)	Empirisch: qualitativ: Längsschnitt-Interventionsstudie	./.	Zugang zum Teilchenkonzept durch Phänomene	Deutschland
3	1	Bindernagel und Eilks (2009)	Empirisch: qualitativ: Interviewstudie	./., betrachtet nur Teilchen	Fachdidaktisches Wissen zu Teilchen	Deutschland
4	1	Bühler und Erb (2010)	Empirisch: quantitativ: single choice aus 4-5 Vorgaben	./.	Metabetrachtung zum Mikrokosmos in Abhängigkeit vom physikalischen Weltbild	Deutschland
5	1	Christen (1994)	Theoretisch: Unterrichtspraxis	- In der Sek I und im Grundkurs besteht der „Stellenwert von Atom- und Bindungsmodellen vielmehr darin, daß sie Werkzeuge darstellen, Hilfsmittel zur Deutung, zum Verständnis von Phänomenen; sie sind also keinesfalls Selbstzweck.“ (S. 30) - D. h. nicht die historische Entwicklung ist entscheidend, sondern der Nutzen des Modells. Erst bei Fragestellungen, die sich nicht anders erschließen lassen, bietet sich ein Modellwechsel an.	Betrachtung und tendenzielle Ablehnung des Orbital-Atommodells.	Deutschland
6	1	Gad und Mittelsten Scheid (2008)	Theoretisch: Unterrichtspraxis	./.	Unterrichtspraktische Aufarbeitung der Seeberge in Form eines Planspiels	Deutschland
7	1	Graulich et al. (2010)	Theoretisch: Unterrichtspraxis	./.	Historisch begründeter Versuch eines unterrichtspraktischen Umgangs mit heuristisch-intuitivem Umgang mit perizyklischen Reaktionen	Deutschland
8	1	Gröger (2010)	Theoretisch: Unterrichtsgeschichte	./.	Historische Genese der Minimumtonne	Deutschland
9	1	Hammer (1988)	Theoretisch: Unterrichtspraxis	./.	Spezifische Betrachtung der Koordinationsverbindungen	Deutschland
10	1	Hilgers et al. (1994)	Theoretisch: Unterrichtspraxis	./.	Spezifische Betrachtung des Versuchsmodells Kronenether als Carriermodell	Deutschland
11	1	Keune und Daemmgen (1988)	Theoretisch: Unterrichtspraxis	./.	Spezifische Betrachtung der verschiedenen Darstellungsmöglichkeiten der Glucose	Deutschland
12	1	Klinger (2001)	Theoretisch: Unterrichtspraxis	- „Modelle stellen dabei eine Abbildung der meist sehr komplexen Wirklichkeit auf einfacher strukturierte Gebilde dar, die die wesentlichen Strukturelemente und Eigenschaften des Originals wiedergeben. Modelle dienen damit der Veranschaulichung eines Sachverhaltes. Sie vereinfachen ihn und machen ihn gedanklich fassbar, so dass er auch einer mathemati-	Metabetrachtung von physikalischen Modellen hinsichtlich der Anschaulichkeit	Deutschland

				<p>schen Behandlung leichter zugänglich ist.“ (S. 167)</p> <ul style="list-style-type: none"> - „Es wird die Existenz einer ‚äußeren Wirklichkeit‘ angenommen, mit der der Mensch in Wechselwirkung tritt. Aufgrund der Informationen, die er über seine Sinne bzw. über Messapparaturen von der äußeren Wirklichkeit erhält, bildet er sich eine sogenannte ‚innere Wirklichkeit‘, also ein Abbild der äußeren Wirklichkeit.“ (S. 168) - Objektbereich und Modellbereich sind nicht immer deckungsgleich. (S. 168) - Modelle müssen immer wieder neu an Entdeckungen der Wissenschaft angepasst oder sogar verworfen werden. (S. 169) - Modelle brauchen in der Physik Analogiebildungen = „Übertragung von Strukturen oder Verfahrensweisen von einem Gegenstandsbereich auf einen anderen“, eine Art „Mustererkennungprozess“. (S. 170) - „Der Wasserstromkreis und der elektrische Stromkreis spielten weder bei Ohm noch bei Fick eine Rolle.“ (S. 172) - Durch „ikonische Darstellungen“ (S. 173) werden komplexe Sachverhalte mitunter erst erschließbar für den Geist. - Funktionsmodelle zur Klärung und Elementarisierung von Abläufen - Ähnlichkeitsmodell als Verkleinerung der Wirklichkeit (S. 183) - Analogmodell als Link zu bereits Bekanntem - Gegenständlich-makroskopische Modelle von mikroskopischen Vorgängen und Strukturen (S. 184) 		
13	1	König und Reiners (2004)	Empirisch: quantitativ: Interventionsstudie	./.	Spezifische Betrachtung der Auswirkung von 3D-Atomdarstellungen (hinsichtlich Bindungen) auf den Lernerfolg	Deutschland
14	1	Leisner und Mikelskis (2004)	Empirisch: quantitativ	- „Ein Lerner hat metakonzeptuelle Kompetenz, wenn er aufgrund von Kenntnissen und erlernbaren Fähigkeiten in der Lage ist, eigenständig und		

				<p>situationsunabhängig:</p> <ul style="list-style-type: none"> - zwischen Phänomenen und Modellen zu trennen, - Modelle zum Problemlösen auszuwählen, einzusetzen und zu bewerten, - über das Problemlösen mit Modellen und das Modellieren zu reflektieren.“ (S. 121) 		
15	1	Lethmate und Arning (2003)	Theoretisch: Unterrichtspraxis	<p>Geo-„Modelle sind vereinfachte Repräsentanten (Abbilder) realer Systeme. Sie entsprechen in wesentlichen Eigenschaften dem Original, sind aber anschaulicher „weil bei der Modellbildung einzelne Merkmale eines komplexeren Systems hervorgehoben bzw. eliminiert werden“ (Killermann 1995, S. 215) (S. 37)</p> <ul style="list-style-type: none"> - Auch Kombinationen, also Modellexperimente, sind in der Geo möglich. (S. 37) - „Experiment und Modell schließen sich also gegenseitig nicht aus, sie liegen vielmehr auf unterschiedlichen Betrachtungsebenen, zum einen auf der Ebene der betrachteten Objekte (Modell oder Original/Natur) und zum anderen auf der Ebene der Arbeitsweisen.“ (S. 37) - Modellexperimente sind wegen Unmöglichkeit, sie am Realobjekt auszuführen, in Geo weit verbreitet. (S. 37) - Nach Claßen (1997) = Mischung aus konkretem und theoretischem Modell; - Problem: Wie adäquat hat das Modell und das Modellexperiment mit der Wirklichkeit zu tun? (S. 37). - Auf Unterschiede sollte hingewiesen werden. (S. 38) 	Abgrenzung der verschiedenen Arbeitsweisen der Naturwissenschaften vor dem Hintergrund des Geografieunterrichts	Deutschland
16	1	Litsche (1985)	Theoretisch: Unterrichtspraxis	<ul style="list-style-type: none"> - Zwei Modellarten sind gesichert: die „theoretischen (geistigen, ideellen, gedanklichen) und die technischen („wirklichen“, materiellen, praktischen) Modelle... Theoretische Modelle existieren nur in Gedanken, die mittels sprachlicher oder anderer Zeichen ausgedrückt werden, technische Modelle existieren in irgendeiner materiellen Form, mit der man 	Metabetrachtung von Modellen hinsichtlich der Unterscheidung von Gedanken- und technischen Modellen, auch hinsichtlich der Unterscheidung der technischen Modelle nach ikonischen („Realmodelle“) und nichtikonischen (Bezug zu Theorie) Modellen	Deutschland

			<p>„manipulieren‘ kann.“ (S. 262)</p> <ul style="list-style-type: none">- Aus der Kybernetik stammt die Vorstellung des „internen Modells der Außenwelt“ auf gedankliche Abbilder übertragen und somit deren „Modellcharakter“ hervorgehoben; beim Gedankenmodell werden „Veränderungen in der Wirklichkeit im ideellen Modell ‚nachgespielt‘“ (S. 262)- Auf der Grundlage des Gedankenmodells können Vorhersagen für die Realität, aber auch umgekehrt versucht werden, die Beobachtung der Realität mit dem Gedankenmodell zu beschreiben – es gibt nicht nur die Realität, sondern auch die „Welt der gedachten Dinge“, und diese Einsicht muss den SuS bewusst sein. (S. 262)- Gedankenmodelle sind keine Anschauungs-, sondern Denkmittel, „Operationen am Modell stehen für Operationen mit Gedanken, nicht für Operationen mit der Realität.“ (S. 263)- „Technische Modelle dieser Art sind materielle Realisierungen gedanklicher Modelle“ und „dienen dazu, Gedanken über einen Bereich der Wirklichkeit auszudrücken“. (S. 263)- „Die Simulation geistiger Vorgänge durch materielle Vorgänge ist ein wichtiges Mittel zur Erhöhung der Produktivität der geistigen Arbeit.“ (S. 263, in: Klaus, G.: Spieltheorie in philosophischer Sicht. Berlin 1968, S. 22)- Sogenannte technische Modelle sollten zur Entwicklung des Denkens genutzt werden, was häufig leider vergessen wird. Die Beziehung des Modells zum Original lässt Litsche in diesem Artikel bewusst ausgeblendet. (S. 263)- „Ein Modell steht immer für etwas anderes, für das Modellerte, wobei der Grad und die Art der Übereinstimmung des Modells mit dem Modellierten sehr verschieden sein können.“ (S. 263)- Unterscheidung der technischen Modelle danach, ob sie für die unmit-		
--	--	--	---	--	--

				<p>telbare Wirklichkeit oder über die Theorie über die Wirklichkeit sind: Bei Modellen zur Theorie über die Wirklichkeit („nicht-ikonisch“) ergeben sich am ehesten Bezugspunkte zur Chemie und Physik (DNA-Modell/Atommodell), bei Realmodellen wie Blütenmodellen (ikonisch) eher nicht. (S. 263)</p> <p>- „Das didaktisch-methodische Problem entsteht dann, wenn nichtikonische Modelle zur Wissensvermittlung benutzt werden. Dabei muß gewährleistet werden, daß diese nicht so eingesetzt werden wie ikonische Modelle, denn wenn das geschieht, dann meinen die Schüler, die Gegenstände der wirklichen Welt sähen so aus wie die angeschauten Modelle.“ (S. 263)</p>		
17	1	Markert (1988)	Theoretisch: Unterrichtspraxis	./.	Spezifische Betrachtung eines Synapsenmodells und der Wirkungsweise von Nervengiften wie E605	Deutschland
18	1	Meyer (1990)	Theoretisch: Unterrichtspraxis	<p>- „Die Benutzung vereinfachender Denkmodelle, die in simplifizierender Weise ein partielles Verständnis ermöglichen, ist ein notwendiger Trick des menschlichen Denksystems, um mit Sachverhalten fertigzuwerden, die seine Kapazität eigentlich überschreiten. ... Wenn das Bewußtsein über einen bestimmten Vorrat an Denkmodellen verfügt, so versucht es, auch neu aufkommende Erfahrungen mit Hilfe dieser schon verfügbaren Denkmodelle zu verstehen (Steinbuch 1977, S. 10/11).“ (S. 4)</p> <p>- „Etymologisch gesehen stellt das Wort «Modell» die Verkleinerungsform von modus, nämlich modulus, dar und bedeutet übersetzt soviel wie Maß/Maßstab oder Art und Weise (vgl. Eschenhagen, Kattmann & Rodi 1985, S. 354).“ (S. 4)</p> <p>- Ebenfalls von Eschenhagen/Kattmann/Rodi (1985) stammt die Abwandlung von Steinbuch (1977) der Darstellung des Modells als ein Original mit „unendlich vielen Eigenschaf-</p>	Metabetrachtung von Modellen im Biologie(unterricht)	Deutschland

				<p>ten“ das auf einer „gegenständlichen Realität“ basiert, dann aus einer Modellierung, also einer „theoriebezogene[n] Reduktion auf das Wesentliche“, die zu einem ideellen Denkmodell führt, welches wesentliche Eigenschaften auf der Bewusstseinssebene präsentiert. Aus diesem entwickelt sich das materielle Anschauungsmodell mit (un)wesentlichen Eigenschaften auf einer gegenständlichen Realitätsebene.</p> <p>– „jede Hypothese in der Biologie ein Denkmodell (vgl. Nachtigall 1972, S. 155)“ (S. 4)</p> <p>- Analogien natürlicher Systeme „struktureller oder funktionaler Art“ sollten als Modelle bezeichnet werden (Halbach 1974, S. 293). (S. 4)</p> <p>- „Modelle sind vereinfachte ideelle oder materielle Abbildungen der Wirklichkeit als Ganzes, eines Ausschnittes oder bestimmter Zusammenhänge der Wirklichkeit, die der Veranschaulichung wesentlicher Struktur- oder Funktionsmerkmale originaler Objekte oder Vorgänge dienen.“ (S. 4)</p> <p>– „Modelle heben bestimmte Merkmale bzw. Funktionen besonders hervor und vernachlässigen andere (didaktische Reduktion).“ (S. 4)</p> <p>- „Modelle sind Hilfsmittel zur Erkenntnisgewinnung bzw. Erkenntnisvermittlung und werden immer dann benötigt, wenn die direkte Untersuchung der originalen Struktur, Funktion oder des originalen Vorgangs unmöglich, zu komplex oder zu zeitaufwendig ist.“ (S. 4)</p> <p>- Bei materiellen Modellen „unterscheidet man Struktur- und Funktionsmodelle.“ (S. 4)</p> <p>- „Gedankenmodelle und Modellvorstellungen existieren in ihrer Reinform nur als Idee, als Gedanke.“ (S. 5)</p> <p>- Auch wenn man Schemazeichnungen daraus macht, gelten sie weiter als Gedankenmodelle. Deswegen sind Gedankenmodelle die häufigsten Modelle im Biunterricht. (S. 5)</p>		
--	--	--	--	--	--	--

			<ul style="list-style-type: none">- Gedankenmodelle sind „-sprachlich formulierte Modellvorstellungen, die auf Beobachtungen oder Experimente zurückgehen,- bildhafte Vergleiche,- mathematische Gleichungen, Gesetze und deren Abbildungen,- chemische Reaktionsgleichungen,- alle vereinfachten bildlichen Darstellungen von chemischen Stoffen durch Symbole von physiologischen (allgemein) oder molekularbiologischen Prozessen durch schematische Darstellungen von Regelungsvorgängen durch Abfolgediagramme oder Pfeildiagramme von Systemen oder Kreislaufprozessen, von Nahrungsbeziehungen von mikroskopischen Originalen durch Schemazeichnungen, von Stammbäumen.“ (S. 5)- „Strukturmodelle veranschaulichen Baumerkmale lebender Organismen; sie sind vereinfachte Nachbildungen morphologischer und anatomischer Merkmale“ (S. 5 f.), aufgeteilt nach Zerlegbarkeit (ja/nein) und Flach- und 3D-Modell (S. 5 f.)- „Funktionsmodelle veranschaulichen das Prinzip von Vorgängen oder Funktionen. Meist handelt es sich um Analogmodelle, deren charakteristische Merkmale sich bestimmten Kennzeichen des Originals oder des originalen Vorgangs zuordnen lassen.“ (S. 6)- Die Ähnlichkeit zum Original ist demnach nicht notgedrungen groß bis nicht vorhanden. „Je größer die Abstraktion, desto stärker tritt das Modellhafte hervor.“ Auch bei Funktionsmodellen kann unterschieden werden zwischen Flach- und 3D-Modellen (S. 6)- Die Modellbildung umfasst „1. die Sammlung und Auswertung von Daten, 2. die theoretische Konstruktion einer Modellvorstellung, 3. die Formulierung der Modellvorstellung,	
--	--	--	--	--

			<p>4. ihre apparative Konstruktion, 5. die schematische Darstellung des apparativen Modells.“ (S. 6)</p> <ul style="list-style-type: none"> - 1-3 ist ähnlich wie bei Wissenschaft, 4-5 in Wissenschaft nicht unbedingt notwendig, aber in Schule wichtig: Bei Wissenschaft heißt es Forschungs-, bei SuS Erkundungsmodell. - Auch die Erwähnung der Grenzen (von Hentig 1971, S. 23) ist wichtig. Es geht um Anschauungsbildung (Memmert 1975, S. 71). (S. 6) - „Das Bilden von Anschauungen ist ein höchst aktiver Prozeß.“ (S. 6) Schule sollte den gesamten Modellbildungsprozess beibringen. (S. 6) - Modelle sind v. a. wichtig für SuS mit Lernschwierigkeiten bei komplexen Zusammenhängen (S. 6 f.) - „Schietzel/Kalipke (1968) sprechen in diesem Zusammenhang davon, daß Modelle ‚antiabstrakt‘ sind.“ (S. 7) Das widerspricht heutiger Vorstellung. - SuS „sind im allgemeinen am meisten von solchen Modellen beeindruckt, die sie in die Hand nehmen, verändern und mit denen sie praktisch arbeiten können. Am intensivsten beschäftigen sich Schüler mit Modellen, die sie selbst herstellen.“ (S. 7) „Untersuchungen (vgl. z. B. Staeck 1980) belegen, daß Schüler bei einem medial vermittelten Unterricht eine wesentlich höhere Behaltensleistung zeigen als eine Kontrollgruppe, die nur verbal unterrichtet wurde.“ (S. 7 bzw. Abb. 4, S. 8) - Es sollte „die Vorläufigkeit mancher Modelle bewußt gemacht werden.“ – „Die Modellkritik sollte grundsätzlicher Bestandteil jeglichen Modelleinsatzes sein.“ (S.8) - „Alle Modelle (müssen) vereinfachen. Der Grad der Reduktion hängt von der jeweiligen wissenschaftlichen Sachstruktur und der Adressatengruppe ab (vgl. Grüner 1967, Weber 1976). Voraussetzung für eine Reduktion ist, daß der zu veranschaulichende Sachverhalt auf eine Kernaussage verdichtet 		
--	--	--	--	--	--

				<p>werden kann." (S. 9)</p> <ul style="list-style-type: none"> - „Eine konsequente Farbgebung hilft, einander entsprechende Teile zu kennzeichnen.“ (S. 9) - „Kuschmann 1971, Siedentop 1971, Staeck 1987: - Modelle sollten die Ideen, Strukturen, Funktionen oder Vorgänge besonders hervorheben, die mit dem Lernziel übereinstimmen. Modelle werden für Lernende in ihrer spezifischen Entwicklungssituation und ihrem spezifischen Vorstellungsvermögen erstellt (Hervorhebungsmerkmal, Subjektivierungsmerkmal). - Die jeweilige Idee, Struktur, Funktion oder Vorgänge müssen sachlich richtig abgebildet sein. - Modelle sollten einfach und anschaulich sein. Irrelevante Zutaten sollten so weit wie möglich vermieden werden, um das Vorstellungsvermögen der Schüler nicht zu überfordern (Einfachheit und Anschaulichkeit).“ (S. 10) 		
19	1	Mikelskis-Seifert (2004)	Empirisch: quantitativ	<ul style="list-style-type: none"> - „Der Modellbegriff wird aus erkenntnistheoretischer und aus lernpsychologischer Perspektive expliziert.“ (S. 15) - Kircher (1995. S. 96): Modell-Objekt-Beziehung als Analogie, Lernpsychologie in Modell-Subjekt-Beziehung berücksichtigen (S. 16) - Leitideen für Unterricht mit explizit-metakonzeptueller Betrachtung der Grundlagen erkenntnistheoretischer Reflexionen: - „Diskussion und Reflexion über die Natur der Modelle und den Modellierungsprozess; - Bewusstes Konstruieren von Modellen für die Beschreibung und Deutung physikalischer Phänomene; - Prüfen der Modellannahmen auf Tragfähigkeit bzw. Aufzeigen der Modellgrenzen; systematisches Trennen von Erfahrungs- und Modellwelt.“ (S. 16) - Beachte das System multipler Repräsentationen in Abb. 3 (siehe Mikelskis-Seifert 2002, S. 105) (S. 17) 	Metabetrachtung von Modellen im Chemie- und Physikunterricht, fokussiert auf den Teilchenmodellbegriff	Deutschland

				- darin Trennung von Erfahrungs- und Modellwelt (S. 17)		
20	1	Mikelskis-Seifert und Leisner (2004)	Empirisch: quantitativ	- „Notwendigkeit der Entwicklung einer Modellkompetenz durch einen Unterricht, der die Modellproblematik ausführlich thematisiert, im Anfangsbereich.“ (S. 123)	Metabetrachtung von Modellen in Physikunterricht umgesetzt (Modellmethode)	Deutschland
21	1	Mikelskis-Seifert et al. (2005)	Theoretisch: auf Unterrichtspraxis ausgerichtet	- Es ist „praktisch unmöglich, ... alle Facetten des Modellbegriffs in unserer Kultur aufzuzeigen.“ (S. 32) - „Ein Modell ist ein Gegenstand oder theoretisches Konstrukt, das von einem Subjekt für einen entsprechenden Zweck geschaffen bzw. verwendet wird. Dabei bestehen zwischen bestimmten Entitäten des Modells und bestimmten Entitäten des präsentierten Objektes Analogien.“ (S. 35) - „In Diskussionen um die mentalen Modelle sind ebenfalls verschiedene Begriffsbestimmungen zu finden. Beispielsweise wird bei Anderson (1996) der Begriff ‚mentales Modell‘ anders verwandt als bei Paivio (1978), bei Mayer (1994) oder auch bei Schnotz (1994, 1998, 1999). Allgemein können zwei Richtungen identifiziert werden. Zum einen wird das mentale Modell als eine bildhafte Repräsentation neben der propositionalen Darstellung von Wissen im Gehirn angesehen (z. B. bei Schnotz bzw. bei Mayer). Zum anderen wird der Begriff des mentalen Modells viel weiter gefasst. Bei den Vertretern dieser Richtung werden mentale Modelle als interne semantische Repräsentationen betrachtet, die der Mensch im Rahmen des Wahrnehmens und des Denkens erzeugt und verwendet (siehe Anderson).“ - In Abb. 4 (S. 33) mentales Modell als Produkt einer didaktischen Rekonstruktion der Realität und als Mittler zum eigentlichen, externalisierten Modell in der abstrakten Welt. (S. 33) - In Abb. 5 (S. 33) werden Modelle nach ihren Zielen in Entwickeln bzw. Anwenden von Modellen unterteilt.	Metabetrachtung von Modellen im Physikunterricht	Deutschland

				- So ist „Analogie als eine partiell isomorphe Abbildung zu verstehen. Das heißt, dass es einerseits Eigenschaften des Objekts gibt, die das Modell nicht besitzt, sonst wäre es ja nicht das Modell. Andererseits kann man einem Modell auch Eigenschaften zuschreiben, die das Objekt nicht aufweist. Beispielsweise besitzen die Kugeln in Modelldarstellungen von Atomen oder Molekülen Form und Farbe, die nicht auf die wirklichen kleinsten Teilchen übertragen werden können.“ (S. 35)		
22	1	Nocke (2001)	Theoretisch: Unterrichtspraxisbeispiel	./.	Spezifische Betrachtung eines Gegenstromspielmodells	Deutschland
23	1	Rueckl und Ebinghaus (1989)	Theoretisch: Unterrichtspraxis	./.	Spezifische Betrachtung eines Feldenergiekonzeptmodells	Deutschland
24	1	Schneider (1981)	Theoretisch: Unterrichtspraxis	- „Definitionen des Modellbegriffs heben im allgemeinen den abstrahierenden Charakter dieser Medien hervor.“ (S. 259) - „Nachtigall (1972) nennt Modelle ‚vereinfachende Abstraktionen von biologischen Substraten‘.“ (S. 259) - anatomisches Modell: „plastische, meist vergrößerte Nachbildungen der Wirklichkeit, die durch Weglassung anatomischer Einzelheiten Wesentliches deutlich hervortreten lassen“ (Killermann 1974) (z. B. Modelle von Auge, Ohr, Zahn) (S. 259) - Funktionsmodelle: „die neben der Struktur auch Bewegungen‘ und Funktionen zeigen (Killermann 1974)“ und im allgemeinen morphologisch stärker abstrahiert, hinsichtlich einer bestimmten Funktion aber realer als anatomische Modelle“ sind. (S. 259) - Realnachbildungen sind „solche Medien, die ein Realobjekt naturgetreu in Größe und Farbe, im allgemeinen aber nicht in ihrer Materialbeschaffenheit nachahmen... und lassen sich in die Nähe der statischen anatomischen Modelle einordnen.“ (S. 259) - Welche Art des Modells ist abhängig vom didaktischen Zweck. (S. 259)	Spezifische Betrachtung verschiedener quantifizierender Wirbelsäulenmodelle	Deutschland

25	1	Steinbeck und Markert (1989)	Theoretisch: Unterrichtspraxis	./.	Spezifische Betrachtung einer Einführung des Regelkreises mithilfe einer Kolbenprober-Apparatur	Deutschland
26	1	Streller und Bolte (2008)	Theoretisch: Unterrichtspraxis	./.	Spezifische Betrachtung einer experimentellen Einführung in die Entstehung des Tornados	Deutschland
27	1	Touché (1989)	Theoretisch: Unterrichtspraktische Hinweise	./.	Spezifische Betrachtung einer experimentellen Einführung in die Entstehung des Tornados	Deutschland
28	1	Winnenburg (2000)	Theoretisch: Unterrichtspraktische Hinweise	./.	Spezifische Betrachtung der Behandlung des Themas Sonnenfinsternis	Deutschland
29	1	Cohen und Yarden (2009)	Empirisch: qualitativ und quantitativ	./.	Spezifische Betrachtung der Curriculumsänderungsumsetzung: Zelle langfristig und begleitend zu anderen Themen einsetzen	Israel
30	1	Crawford und Cullin (2004)	Empirisch: qualitativ	<p>- "A model of an object or a phenomenon is a simplified imitation that, hopefully, helps our understanding" (AAAS 1989, 1993). (S. 1381)</p> <p>- "models represent <i>ideas</i> and not the physical objects." (S. 1382)</p> <p>- "1. Models exist as aids to understanding phenomena, and this understanding can be checked or verified by comparing the results obtained by manipulating the model to observations obtained in the real world.</p> <p>2. A primary guideline for making a model is to consider its purpose.</p> <p>3. A scientist can have more than one model for the same thing because different models can be used to address different specific interests or questions about the referent.</p> <p>4. A scientific model can change and be replaced by one that is better for answering questions." (S. 1382 f.)</p> <p>- Forscher fragen nicht nach der Richtigkeit von Modellen, sondern eher: „(1) Can the model <i>explain all the observations</i>? (2) How can the model be used to <i>predict</i> the behavior of the system if it is manipulated in a specific way? (3) How is the model <i>consistent with other ideas</i> about how the world works and with other models in science?“ (S. 1382)</p>	Generelle/Metabetrachtung von Modellen, insbesondere von dynamischen Modellen, zu erstellen mit Model-IT	USA

31	1	Drechsler und van Driel (2008)	Empirisch: qualitativ	- "Models link theories with a target – a system, an object, a phenomenon, or a process; they are parts of theories scientists develop to explain some aspects of the world-as-experienced (Gilbert et al. 2000)." (S. 612)	Spezifische Betrachtung der Vermittlung des Themas Säuren und Laugen und welche Modelle dahinter stehen	Schweden (Sample), in Kooperation mit den Niederlanden
32	1	Eilks et al. (2009a)	Theoretisch: auf Unterricht ausgerichtet		Spezifische Betrachtung von internetgestützten Visualisierungen und Animationen zu Daniells voltaischer Zelle und Diskussion der Für und Wider	Deutschland
33	1	Garritz (2013)	Theoretisch: auf Unterricht ausgerichtet: Plädoyer für Historisierung und Philosophisierung, um adäquate NoS-Vorstellungen (Philosophie der Naturwissenschaften) zu erzeugen	./.	Spezifische Betrachtung der historischen und philosophischen Kontroversen und deren Vermittlung bzgl. des Themas Quantenmechanik/-chemie	Mexiko
34	1	Harrison und Jong (2005)	Empirisch: qualitativ	- "Teachers' analogies exhibit a rich variety of form and content (Dagher, 1995) and teacher analogies can be planned or spontaneous (Thiele & Treagust, 1994). Successful analogies are systematic, include multiple mappings and utilise relational thinking (Gentner & Medina, 1998). Pictorial and role-play analogies are frequently used to enhance analogical familiarity but few teachers discuss analogical limitations." (S. "1")	Spezifische Betrachtung des Analogieeinsatzes eines Lehrers zu Reaktionsgleichgewicht usw.	Australien, Niederlande
35	1	Jong und van Driel (2004)	Empirisch: qualitativ	./.	Spezifische Betrachtung von PCK-Entwicklung hinsichtlich der Betrachtungsmöglichkeiten verschiedener Chemiethemen (submikroskopisch, makroskopisch, symbolisch): "(...) naturalistic case study of the development of eight student teachers' pedagogical content knowledge (PCK) of the multiple meanings of chemistry topics ... interviewing the student teachers individually before and after the lessons." (S. 477) (Es geht um Studierende, die bereits M.Sc. haben und die Lehrbefähigung erwerben wollen.)	Niederlande
36	1	Jong et al. (2005b)	Empirisch: qualitativ	- "representations (e.g., models, metaphors" (S. 949) - "Generally speaking, a model in science may be defined as a non-unique, partial representation of a target, focusing on specific aspects of	Spezifische Betrachtung von Teilchenkonzept: "obtained from answers to written assignments, transcripts of workshop discussions, and reflective lesson reports, written by the participants."	Niederlande

				<p>it, whereas other aspects of the target are deliberately excluded (Ingham & Gilbert, 1991).” (S. 950)</p> <p>- “In the second half of the last century, the production and use of models played a central role in the growth of chemical knowledge (Luisi & Thomas, 1990). Thinking and reasoning with models enables chemists to visualize the abstract processes and entities they are investigating (Justi & Gilbert, 2002).” (S. 950)</p> <p>- “Chemists often use models without being aware of it. They may, for instance, ‘jump’ from the world of corpuscular entities to the level of macroscopic phenomena, and back, in a flexible and implicit way (Johnstone, 1993). Although this may not be problematic in the context of communication between chemists, it may easily lead to misunderstanding in the context of chemistry education.” (S. 950)</p> <p>- “For secondary students, the conceptual demands of switching between models and phenomena can be overwhelming (Andersson, 1990).” (S. 950)</p>	(S. 947)	
37	1	Justi und van Driel (2005a)	Empirisch: qualitativ	<p>-“a model may be defined as a non-unique, partial representation of a target, focusing on specific aspects of it (Gilbert, Boulter, & Elmer, 2000).” (S. 199)</p> <p>- “‘target’ refers to, for instance, a system, an object, an event, a process, or an idea.” (S. 199)</p> <p>-Modelle sind “main products of science (Gilbert, Boulter, & Rutherford, 1998).” (S. 200)</p>	Spezifische Betrachtung der Entwicklung des Wissens ums Modellieren und von Modellen in der Schule einer bereits graduierten Lehrerin in einem action research-Ansatz	Niederlande
38	1	Justi und van Driel (2005c)	Empirisch: qualitativ	<p>- “Representation is one of the major roles of models in science (Giere 1999, Morrison and Morgan 1999, Nersessian 1999).” (S. 549)</p> <p>- Modeling = “One of the most important parts of scientific reasoning is thus carried out during the dynamic process of producing and revising models.” (S. 549)</p>	Spezifische Betrachtung der Wirksamkeit (Intervention) eines Ausbildungsprogramms hinsichtlich CK, PCK und Curriculum Knowledge beim Modellieren	Niederlande

39	1	Justi und van Driel (2005b)	Empirisch: qualitativ	Siehe 38 (identisch)	Spezifische Betrachtung der Wirksamkeit (Intervention) eines Ausbildungsprogramms hinsichtlich CK, PCK und Curriculum Knowledge beim Modellieren	Niederlande
40	1	Khourey-Bowers und Fenk (2009)	Empirisch: quantitativ	<p>- "Scientific models allow learners to explain and generalize abstract phenomena by using a schema or network, which integrates discrete concepts into a complex system of interrelated constructs. Models provide visual or verbal expressions which transform personal knowledge into public knowledge and serve as one means of initiating learners into the scientific discipline." (S. 442)</p> <p>- "Chemistry models include representations at macroscopic, microscopic or particulate (referred after this as particulate), and symbolic levels. Representations at the macroscopic level focus on observable properties and processes and at the particulate level focus on arrangement and motions of particles (Bunce and Gabel 2002). Symbolic models focus on chemical and mathematical notations and equations (Harrison and Treagust 2000; Kruse and Roehrig 2005)." (S. 442)</p>	Spezifische Betrachtung der Wirksamkeit (Intervention) eines Ausbildungsprogramms hinsichtlich CK, PCK und Selbstwirksamkeits-erwartung beim Unterrichten mit Modellen	USA
41	1	Liang et al. (2011)	Empirisch: quantitativ	./.	Spezifische Betrachtung der Gasteilchen in Abhängigkeit des Drucks bei Acht- und Neuntklässlern (und Grundschullehrkräften)	Taiwan
42	1	Lin und Chiu (2010)	Empirisch: qualitativ	./.	Spezifische Betrachtung der mentalen Modelle/Präkonzepte von drei guten und drei schlechten SuS zu Säuren und Basen und wie eine Lehrerin darauf reagiert (inadäquat bei schlechten SuS)	Taiwan
43	1	Nelson und Davis (2012)	Empirisch: qualitativ	- "Scientific models are two- or three-dimensional representations that highlight the central features and key relationships among components of a simplified system or scientific phenomena for the purposes of understanding, communicating, and/or generating predictions about the system or phenomena in question (e.g., Gilbert & Boulter, 2001; Harrison &	Spezifische Betrachtung der Wirksamkeit einer Fortbildungsmaßnahme bei Grundschullehrkräften hinsichtlich PCK zu Modellen	USA

				<p>Treagust, 2000)." (S. 1932)</p> <p>- „Modeling practice is a nonlinear, iterative approach to learning science content, in which students take an active, evidence-based role in reshaping their own conceptual understandings of the science content (Schwarz et al., 2009)." (S. 1932)</p> <p>- "Understandings about the nature and purposes of scientific models and modeling practice, collectively termed 'metamodeling knowledge' (Schwarz & White, 2005), ground the critique of student-generated scientific models." (S. 1933)</p> <p>- "model quality, including model clarity, salience, and consistency with empirical evidence (Schwarz et al., 2009)." (S. 1933)</p>		
44	1	Oh und Kim (2013)	Empirisch: qualitativ	./.	Spezifische Betrachtung von Unterrichtsvideomitschnitten und dem Einsatz von Repräsentationen im Grundschulunterricht	Südkorea
45	1	Okanlawon (2010)	Theoretisch: Rahmen	./.	Spezifische Betrachtung der Stöchiometrie als Teil der Modellkompetenz einer Lehrkraft	Nigeria
46	1	Padilla et al. (2008)	Empirisch: qualitativ	./.	Spezifische Betrachtung des Konzepts von Chemieprofessoren zur Stoffmenge, basierend auf Mortimers Conceptual Profile Model (CPM)	Mexiko/Argentinien
47	1	Papageorgiou et al. (2010)	Empirisch: quantitativ	./.	Spezifische Betrachtung der Wirksamkeit einer Fortbildung zu Aggregatzuständen	Griechenland (UK)
48	1	Romine und Walter (2014)	Empirisch: quantitativ	./.	Spezifische Rasch-Betrachtung der kriterialen Validität eines Tests zu Makroevolution (MUM)	USA
49	1	Schmelzing et al. (2013b)	Empirisch: quantitativ	./.	Spezifische Betrachtung eines deklarativen Biologielehrkräftetests zum Blutkreislauf	Deutschland
50	1	Šorgo et al. (2014)	Empirisch: quantitativ	./.	Spezifische länderübergreifende Betrachtung eines selbst entwickelten Tests zur Einstellung zu NoS, Evolution und Genetik	Tschechien, Slowakei, Slowenien, Türkei
51	1	Strübe et al. (2014)	Empirisch: quantitativ	<p>- „they function as a bridge between scientific theory and the world-as-experienced ('reality')" (Gilbert, 2004, p. 1169). (S. 381)</p> <p>- "Justi and Gilbert (2002a, 2003)</p>	Spezifische Betrachtung einer Testentwicklung zum PCK von Lehrkräften hinsichtlich Chemiemodellen (basierend auf Tepner et al.) (S. 19)	Deutschland

				<p>describe the role of models in science education: Students should know the most important models in science, how they were developed and the limitations of models. They should develop and test their own models and know about the importance of models when scientific findings were disseminated and accepted (Justi and Gilbert, 2002a; 2002b; 2003). In this context Gilbert (2004) speaks of 'Learning to Use Models', 'Learning to Revise Models', and 'Learning the Reconstruction of a Model'." (S. 381)</p> <p>- "it is important to discuss the limitations of models (Justi and van Driel, 2005; Saari and Viiri, 2003) and to carve out the change or replacement of models It is also necessary for learning to use different models which represent a concept under different aspects or for different purposes (Grosslight, Unger, and Jay, 1991; Harrison and Treagust, 2000; Saari and Viiri, 2003). The colour of a model can lead to students' misconceptions, because of this it is important to discuss the function of the colour in class (Justi and van Driel, 2005). Teachers should involve students in modelling processes, by creating, developing, building, testing, communicating, and reflecting their own models (Gilbert, 2004; Grosslight et al., 1991; Henze, Van Driel, and Verloop, 2007a; Justi and Gilbert, 2003; Maia and Justi, 2009)." (S. 381)</p>		
52	1	Unal et al. (2014)	Empirisch: qualitativ	<p>- "As seen in the classifications above, mental models always take place within the classifications in literature. They are internal representations having structural similarities in the real life events or processes. By visualizing concepts and processes in the mind, mental models are generated. Mental models are personal, internal and inconsistent with scientific explanations. They develop parallel to the acquisition of new information. It is specific and functional for the person</p>	Spezifische Betrachtung des Wissens von Lehrerausbildern zu Modellen der Naturwissenschaften	Türkei

				who is the owner of the model. As stated by Norman (1983), mental models are mental presentations structured through interaction with reality and different mental models can be formed for one single system. Mental model is an internal process that is structurally similar to events or processes, and it plays the role of calculating personal thoughts for the estimation and explanation of physical phenomenon." (S. 675)		
53	1	Usak et al. (2011)	Empirisch: qualitativ	./.	Spezifische Betrachtung des Wissens von künftigen Chemielehrkräften und -ausbilderInnen zu chemischen Reaktionen	Türkei (Deutschland)
54	1	Valanides und Angeli (2006)		- "A model is an external representation, which can be executed or manipulated by the learner in order to control variables and test hypotheses. A model constitutes a conceptual system and consists of objects or entities, variables or characteristics, and cause- and-effect relationships among variables (Lesh & Doerr, 2003). In essence, a model of a phenomenon constitutes a simplified analog, which does not exactly match in complexity the real one, but it is helpful enough to study and better understand the real phenomenon. Gilbert (1991) suggested that science should be viewed as a process of constructing predictive conceptual models. This will enable students to analyze and synthesize scientific facts, as well as integrate them with scientific theory and give them a unified view of science (Gilbert, 1993; Hestenes, 1987)." (S. 88)		Zypern
55	1	Van Dijk (2009)	Empirisch: qualitativ	./.	Spezifische Betrachtung des PCKs von Biofachleitern im Gymnasium zur Evolution	Deutschland
56	1	J. H. van Driel et al. (2002)	Empirisch: qualitativ	./.	Spezifische Betrachtung der PCK-Entwicklung von künftigen Chemielehrern hinsichtlich mikro – makro	Niederlande: (Utrecht und Leiden)
57	1	Wang et al. (2014)	Empirisch: qualitativ (mit quantitativen Anteilen)	- "A Model is a simplified representation of a system or phenomenon that focuses attention on specific aspects or	Spezifische Betrachtung des Wissens chinesischer Chemielehrer zu Modellen und Abbildungen	China

			components of a system (e.g., prototype), such as ideas, objects, events or processes (Gilbert et al. 1998; Ingham and Gilbert 1991).” (S. 211)		
--	--	--	---	--	--

9.2.2 Tabelle mit zentralen Forschungsergebnissen, Forschungsfragen, Auswertungsmethoden, Stichprobengröße und Population

		Quelle	Zentrale Forschungsergebnisse bzw. Aussagen	Forschungsfrage	Auswertungsmethode	Größe der Stichprobe	Population der Befragten
1	2	Beerenwinkel und Parchmann (2008)	<p>- „Während Experten wie Lehrer/-innen diesen Wechsel leicht und meist unbewusst vollziehen, können die Schüler/-innen den mentalen Sprüngen oft nicht folgen.“ (S. 13)</p> <p>- Vgl. O. de Jong und van Driel: The development of preservice teachers' concerns about teaching chemistry topics at a macro-micro-representational interface. In: Proceedings of the Second International Conference of the European Science Education Research Association "Research in Science Education: Past, Present, And Future" in Kiel, 1999 (S. 13)</p> <p>- „Metadiskussionen über Modelle werden daher als ebenso wichtig für die Entwicklung des Modellverständnisses angesehen wie das intensive und häufige Arbeiten mit Modellen.“ (S. 14)</p> <p>- Vgl. L. Crosslight, C. Unger, E. Loy und C. L. Smith. 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 (1991). (S. 14)</p> <p>- „1. Jedes Modell hatte seine Berechtigung und eignet sich auch heute noch für die Erklärung bestimmter Phänomene. 2. Experimentelle Befunde, die mit dem aktuellen Modell nicht mehr erklärt werden können, führen zu einer Weiterentwicklung des Modells. 3. Jedes Modell hat seine Stärken und seine Grenzen.“</p>	./.	./.	./.	./.

			4. Im Unterricht werden oft Veranschaulichungen von Denkmodellen verwendet.“ (S. 14)				
2	2	Benedict und Bolte (2009)	<ul style="list-style-type: none"> - Der Erklärungsansatz auf der Grundlage eines Teilchenkonzeptes wird „durch unsere Intervention präsenter und/oder plausibler“. (S. 460); - Ein großer Teil von Dritt- und Viertklässlern hat bereits Vorstellungen über Teilchen. - Der Konzeptwechsel zu einem adäquaten Teilchenkonzept ist schwierig. (S. 459) 	<ul style="list-style-type: none"> - „Welche Vorstellungen formulieren Kinder im Grundschulalter überhaupt zur Erklärung von Phänomenen?“ - Verwenden Grundschulkinder von sich aus ein Teilchenmodell? - Wie verändern sich die Erklärungsansätze der Kinder, insbesondere derer, die an unserer Intervention teilnehmen?“ (S. 458) 	videografierte, teilstandardisierte und problemorientierte Einzelinterviews	N = 32 (n = 18 SuS = Interventionsgruppe, n = 14 = Kontrollgruppe)	GrundschülerInnen (8-10 Jahre)
3	2	Bindernagel und Eilks (2009)	<ul style="list-style-type: none"> - Lehrmeinung: Teilchenkonzepte chronologisch in historischer Abfolge unterrichten - Nicht Lernschwierigkeiten stehen im Vordergrund: Unterricht beginnt mit undifferenziertem Teilchen- bzw. Kugelteilchenmodell bzw. Daltonmodell bis hin zu komplexen Atom- und Bindungsmodellen. Roadmap hilft für „Visualisierung“ und „Kommunikation“ (vgl. S. 163). 	Modellverständnis hinsichtlich Teilchenkonzept	Interviewanalyse mit der „Roadmap“, einer Sonderform der Concept-Map	N = 28	Erfahrene Chemielehrkräfte
4	2	Bühler und Erb (2010)	<ul style="list-style-type: none"> - „Die Kinder kommen mit einer aristotelischen Kontinuumstheorie der Materie in die Schule, die sich weithin als resistent gegenüber der im Physik- und Chemie-Unterricht besprochenen Atomtheorie erweist.“ (S. 83) - Vorstellungen über Naturvorgänge lassen sich in „deterministisch/mechanistisches, ... indeterministisches ... und ...theistisches Denken“ (S. 84) aufteilen. - „Jugendliche, die über das Universum als Ganzes nicht deterministisch dachten (das ist praktisch die Hälfte der Befragten), stellten sich die Welt im Kleinen (z. B. fallendes Blatt) streng determiniert vor (mehr als drei Viertel der Befragten). Jugendliche, die den Einfluss Gottes im Großen geltend 	Denken Schülerinnen und Schüler der Klassenstufe 10 (eigentlich 11/12?!) mehrheitlich deterministisch-mechanistisch? Wie sehen sie den Aufbau der Materie? Welche Vorstellungen von Elektronen haben sie?	Prozentuale Antwortverteilung	N = 69	11. und 12. Klasse Gymnasium

			<p>machten, schrieben ihm für den Ablauf kleinerer Naturvorgänge keine Bedeutung zu.“ (S. 84)</p> <ul style="list-style-type: none"> - Unterschiede in deterministischer Betrachtungsweise auf kontinuierlicher und submikroskopischer Ebene - Ca. 60 % der Jugendlichen, die Atomtheorie verwenden, 30 % reine Kontinuums- und 30% Mischvorstellung - „90 % der Jugendlichen waren Anhänger der Teilchentheorie bei Elektronen.“ (S. 85) 				
5	2	Christen (1994)	<ul style="list-style-type: none"> - Auf ein Orbitalmodell kann in Sek I und Grundkursen verzichtet werden, weil nach der Einführung bereits auf die Erklärung der Struktur des Methan-Moleküls (s-/p-/d-Orbital) verzichtet werden muss. - Für den Grundkurs reicht das Kugelwolken-/Elektronenpaarabstoßungsmodell, auch wenn man damit an Grenzen stößt. Diese sind sogar besonders fruchtbar für das Wissenschaftsverständnis. - Im Leistungskurs ist genug Zeit, chronologisch Atommodelle einzuführen, also evtl. auch das Orbitalmodell, allerdings vereinfacht in Welle-, Teilchen- und Kastenmodell. 	./.	./.	./.	./.
6	2	Gad und Mittelsten Scheid (2008)	<ul style="list-style-type: none"> - Die Entstehung des Artenreichtums, aber auch die Gefahren der kommerziellen Nutzung, können an Seebergen gut veranschaulicht werden - Seeberge für 9./10. Klasse nutzbar hinsichtlich Räuber-Beute-Beziehungen in Abhängigkeit von „Umweltfaktoren“, „selektive und adaptive“ (S. 31) Effekte - Aspekte der Evolution und methodisch als Planspiel 	./.	./.	./.	./.
7	2	Graulich et al. (2010)	<p>„Lernenden sollte vermittelt werden, innovativ und kreativ zu sein.“ (S. 449)</p>	./.	./.	./.	./.

8	2	Gröger (2010)	Die historische Analogienfolge bis zur 1910 entstandenen Minimumtonne wird aufgezeigt.	./.	./.	./.	./.
9	2	Hammer (1988)	<ul style="list-style-type: none"> - Koordinationsverbindungen sind keine Moleküle und das ist auch in seiner historischen Entwicklung interessant: - Vorstellungübertragung vom einen zum anderen Gebiet - Historische Modellentwicklung als interessantes Beispiel; eckige Klammern bei Formelschreibweise nach Werner - „Je nach dem Aspekt oder der Absicht, bestimmte Phänomene (etwa Farbigkeit, besondere Stabilität, Reaktionsverhalten oder andere Eigenschaften) zu deuten, empfiehlt es sich jedoch, jeweils verschiedene Modellvorstellungen als Interpretationsmuster heranzuziehen.“ (S. 35) - „Koordinations-einheiten, die mit Buchstaben, Zahlzeichen und anderen zusätzlichen oder ersetzenden Zeichen symbolisiert sind, lassen sich auf eine einleuchtende Weise verstehen und einordnen, wenn sie als abgestuft vereinfachte Wiedergaben eines in einem konkreten Sachmodell repräsentierten Informationsgehaltes gesehen werden.“ (S. 35) - Schrittweise Umformung des räumlichen Molekülmodells hin zur abstrakten eckigen Klammer-Schreibweise nach Werner (Abb. S. 36) - Ligandenabkürzungen (S. 37) - Plädoyer für die Darstellung von konkretem Molekül zu Summenformel (S. 37): - „abnehmende[r] Anschaulichkeit und zunehmende[r] Abstraktion.“ (S. 37), - „fördert es das räumliche Vorstellungsvermögen“ (S. 37) 	./.	./.	./.	./.
10	2	Hilgers et al. (1994)	- Die drei Membrantransportformen Osmose, katalysierte Permeation und aktiver Transport müssen für	./.	./.	./.	./.

			<p>OberstufenschülerInnen erfahrbar gemacht werden. (S. 179)</p> <ul style="list-style-type: none"> - Osmose: kein Problem, bei katalysierter Permeation und aktivem Transport bestehen fast nur Abbildungen zur Vermittlung. (S. 179) - Kronenether sind vergleichsweise günstige Versuchskemikalien. - Biomembranmodelle = Cellulose-triacetat-Folie (S. 179) - „1. Es können qualitative und quantitative Versuche zu einem Thema durchgeführt werden, das ansonsten im Schulunterricht nur theoretisch behandelt wird. 2. Die Versuchsmethode entspricht der in der biologischen Forschung verwendeten Methode zur Untersuchung von Membranen. 3. Die Schüler erfahren an einem konkreten Beispiel, daß die Selektivität eines Carriers nicht nur ein theoretisches Konstrukt ist, sondern tatsächlich Realität besitzt. Der Weg, auf dem diese Erkenntnis gewonnen wird, ist ein typisches Beispiel für die Art der Erkenntnisgewinnung in der Biologie und in den Naturwissenschaften allgemein. 4. Auch die Tatsache, daß bei der Erarbeitung der biologischen Sachverhalte und Aussagen chemische und physikalische Fakten berücksichtigt werden müssen, ist ein wichtiger Aspekt bei der Art der Erkenntnisgewinnung in der Biologie.“ (S. 183) 				
11	2	Keune und Daemmgen (1988)	<ul style="list-style-type: none"> - Kohlenhydrate sind Teil des Bildungsplans der Sek II (S. 8). - Die drei Darstellungsformen des Glucosemoleküls werden zu unterschiedlichen Zwecken benutzt. Oft steht neben den Strukturformeln ein Foto des Molekülmodells. So wird der Fehleindruck vermittelt, das Zickzack-Original würde nur eingeebnet. (S. 9) - Baut man das in die Strukturformel gefasste Wissen wieder mit 	- „Ziel dieser Arbeit ist es, das Zustandekommen der einzelnen Projektionen und ihre Beziehungen untereinander darzustellen.“ (S. 9)	./.	./.	./.

			<p>Molekülbaukästen nach, lassen sich die eigentlich angestrebten Ringe meist nicht schließen. Grund:</p> <p>„(...) 1. so gut wie nie über das Zustandekommen von Projektionsformeln (Struktursymbolen) gesprochen wird,</p> <p>2. nur selten darauf hingewiesen wird, daß die übliche Darstellung der Zucker in Fischer-Schreibweise erfolgt, also eine spezielle (verabredete) Projektionsart darstellt,</p> <p>3. für den Fall, daß der Name FISCHER wähnt wird, diese besonderen Projektionsregeln nicht, falsch oder unzureichend erläutert werden.“ (S. 9)</p>				
12	2	Klinger (2001)	<ul style="list-style-type: none"> - Erst kommen die qualitativen Elemente und Strukturen eines Phänomens, dann erst die mathematische Beschreibung (S. 167), - erst die anschauliche Erfassung einer Struktur, dann die (mathematische) Modellvorstellung. - Analogien in den Begrifflichkeiten Atomkern, -hülle (S. 170) - „Beim Literaturstudium stieß er [Ohm] auf Fouriers Wärmeleitungstheorie und erkannte sofort, dass zwischen der Wärmeleitung und der Elektrizitätsleitung ganz ähnliche, experimentell festgestellte Strukturen vorliegen.“ (S. 171) - „Das Teilchenbild geht von der Vorstellung aus, dass Materie nicht unendlich fein unterteilt werden kann, sondern aus kleinsten, nicht mehr weiter zerlegbaren Partikeln aufgebaut ist.“ (S. 180) 	./.	./.	./.	./.
13	2	König und Reiners (2004)	<ul style="list-style-type: none"> - Computergestützte Darstellung von submikroskopischen Teilchen (S. 334) - Modul 1 die Anwendung des Elektronenwolkenmodells auf Atombau, Modul 2-5 zu verschiedenen Bindungstypen (S. 334) - Klasse 10: „5 Doppelstunden und erfolgte unter Konstanzhaltung der Lehrperson. der Unterrichtsinhalte 	Haben die SuS, denen computergestützte Darstellungen von submikroskopischen Teilchen hinsichtlich Bindungen präsentiert werden, einen besseren Lernerfolg? (selbst formuliert)	Auszählen der Pre-/Postergebnisse, Kontrollgruppe nicht gleiche Klassenstufe, Testwiederholungseffekte nicht kontrolliert	N nicht angegeben (nur: zwei 10. Klassen)	9. und 10. Klasse

			<p>und der Untersuchungsbedingungen. Die Strukturierung variierte durch den alternativen Einsatz der entwickelten Lehr- und Lernhilfen.“ (S. 335)</p> <p>- Klasse 9: Erprobung restlicher Module</p> <p>-,dass die Lernenden mit Computervisualisierung deutlich bessere Ergebnisse bei der Überprüfung ihrer Leistungen in den o.g. Bereichen erzielten als die Lernenden ohne Computerunterstützung.“ (S. 335)</p> <p>- „unter Zuhilfenahme der erstellten Lehr- und Lernhilfen eine angemessenere Vorstellung zur Chemischen Bindung geschaffen werden kann.“ (S. 335)</p>				
14	2	Leisner und Mikelskis (2004)	<p>- Studierende mit „metaconceptual awareness“ (Vosniadou und Ionides (1998)) sind in der Lage zwischen unterschiedlichen physikalischen Konzepten zu unterscheiden. (S. 120)</p> <p>- Es geht dabei um „längerfristige Unterrichtswirksamkeit“ (Fischler und Peuckert (1999)).“ (S. 120)</p> <p>- Konzept wurde weiterverfolgt (Mikelskis-Seifert (2002), ergänzt um „stringente Trennung von Erfahrungswelt und Modellwelt in Anlehnung an Le Maréchal“ (S. 120): Metakzeptentwicklungsmöglichkeit von SuS zu Teilchen empirisch belegt</p> <p>- „Lernen über Modelle“ (Kircher 1995): „Modellbildung, Durchsetzung des Modells, Bewährung des Modells, Reflexion der Modellmethode“. (S. 120)</p> <p>- Gilbert (1997):</p> <p>„1) Modellverständnis entwickelt sich langsam“; 2) es braucht „generelles Wissen über die allgemeine Natur eines Modells, um ein konkretes, ein spezielles Modell verstehen zu können.“ (S. 121)</p> <p>- In jeder Klassenstufe wurden zwei</p>	Haben die SuS nach einem Unterricht mit Modellen mehr Wissen als nach einem traditionellen Unterricht? (selbst formuliert)	Zuordnung der Antworten zu Modellstufen (0-4): 0 = unangemessen, 2 = unsicher, 4 = angemessenes Modellverständnis	Nicht angegeben	Klassenstufen 7-10

			<p>Wissensbereiche abgefragt: „domänenspezifisches Modellwissen und domänenübergreifendes Modellwissen“. (S. 121)</p> <p>- „Im Bereich des domänenübergreifenden Modellwissens findet sich in allen Klassenstufen ein signifikanter Zuwachs, wobei in den Klassen 7 und 8 größere Effekte erzielt werden. Die ausführliche Einführung zum Charakter der Modelle scheint sich auszuzahlen.“ (S. 121)</p> <p>- „Ergebnisse der Untersuchung bestätigen die Forderung nach einem systematischen Unterricht über Modelle.“ (S. 122)</p>				
15	2	Lethmate und Arning (2003)	<p>- Um der Erkenntnisgewinnung zu entsprechen, muss neben dem Experimentieren auch das Untersuchen und Modellieren im Unterricht behandelt werden. (S. 35)</p> <p>- Abgrenzung der Arbeitsweisen Experiment, Untersuchung, Beobachtung, Modell, Modellexperiment, Naturexperiment</p>	./.	./.	./.	./.
16	2	Litsche (1985)	<p>Anhand des Gedankenmodells Zoo mit allen denkbaren Tierarten und der Frage, wie dieser organisiert sein müsste, wird den Fünftklässlern klar, dass es eine spezielle Art der „Beziehungen zwischen den wirklichen Dingen und den Gedanken“ (S. 262) gibt.</p>	./.	./.	./.	./.
17	2	Markert (1988)	<p>- Die Wirkungsweise von Synapsen wird erklärt, auch Nervengifte wie E605, Hinweis auf Drogen und Abhängigkeit der Wirkung von der Dosierung.</p> <p>- „Gefragt nach möglichen Wirkungsweisen, wurden tatsächlich alle vier möglichen Wirkungsweisen des Giftes, wie sie in der Thematik angeführt werden, von den Schülern ohne weitergehende Unterstützung durch den Lehrer genannt.“ (S. 171)</p>	./.	./.	./.	./.
18	2	Meyer (1990)	- Modelle nehmen Ende der	./.	./.	./.	./.

			1980er-Jahre in Publikationen stark zu. (S. 4) - Originale Begegnung besser als Modell (S. 7)				
19	2	Mikelskis-Seifert (2004)	<ul style="list-style-type: none"> - Schülervorstellungsforschung (z. B. Ben-Zvi et al., 1986; Anderson, 1990; Griffiths & Preston, 1992; Duit, 1992; Groslight et al., 1991; Harrison & Treagust, 1996; Treagust et al., 2002) (S. 15) - innovative Unterrichtsvorschläge (Buck, 1981; CLIS, 1987; Séré, 1990, 1992; Meheut & Chornat, 1990; Vollebregt et al., 1997, 1998). (S. 15) - Gute Formulierung: „Ziel einer solchen Analyse ist es, Auskunft über die erfolgversprechenden Elemente der einzelnen Konzeptionen zu erhalten, die bei der Konstruktion eines Unterrichts über Modelle einfließen sollen.“ (S. 15) - Reflexion über Wissenschafts- und Erkenntnistheorie sinnvoll und notwendig (S. 16) - SuS sollen „den artifiziellen Charakter von Modellen erkennen und ein angemessenes Modellverständnis entwickeln“. (S. 16) - „Lernen mit multiplen Repräsentationen“ (Weidenmann, 1997; Sumfleth & Telgenbüscher, 1998, 2000; Schnotz, 1999) - Studien zur Metakognition in dem Ansatz „über Modelle lernen“ (Hasselborn 1998) (S. 16) - zwölf Unterrichtsstunden mit vier Phasen: Einführung (für systematische Trennung von Erfahrungs- und Modellwelt motivieren) Erarbeitung (Volumenzunahme beim Erwärmen von Wasser und Suche nach Gründen als Modellannahme und Überprüfung am Luftkissentisch, im Plenum Präsentation, Diskussion und Einigung auf Teilchenmodell), Anwendung (Übertragung auf anderes Phänomen: Verdunsten in Erfahrungs- und Modellwelt) und 	<ul style="list-style-type: none"> - „Es sollte untersucht werden, ob es den Schülerinnen und Schülern gelingt, stabile“ und adäquate „Metakonzepte zur Teilchenvorstellung zu entwickeln“. (S. 20) - Gibt es „ähnlich positive und signifikante Effekte im Anfangsunterricht der Klasse 8“? (S. 21) - Welche „Bedingungen für die Entwicklung eines metakonzeptuellen Denkens können empirisch“ ermittelt werden? (S. 24) - Das Hauptziel ist es, „die erhobenen Schülervorstellungen und deren Veränderungen mit den beobachteten Unterrichtsmerkmalen in Beziehung zu setzen“. (S. 19) 	<ul style="list-style-type: none"> - Wirksamkeitsanalyse als Pre-Post-Follow-Up-Design (Fragebogen (mehrere Skalen zu verschiedenen physikalischen Vorstellungsbereichen und Problemlöseaufgaben + Fragebogen zu Teilchenvorstellungen) und Concept Map und Audioaufnahmen) (S. 19 f.) - Varianz- und Korrelationsanalysen, Kausalanalysen mithilfe von LISREL. LatentClass-Analysen mit WINMIRA sowie Auswertungen auf der Basis von Modalnetzen sowie Schwerpunkt auf Triangulation 	N = ca. 120 (Klasse 9/10, nicht jedoch für Klasse 8 dargestellt)	9./10., später auch 8. Klasse Gymnasium (Brandenburg)

		<p>Reflexion (Nachdenken, Präsentation von Bildern aus dem Nanokosmos) (S. 17)</p> <ul style="list-style-type: none"> - Ergebnis: meist höchster Wert nach drei Monaten; v. a. bei Modellverständnis sehr große Zunahme (S. 21) - Es fehlt die Vergleichs- und Wartekontrollgruppe ☹. - Es „zeigte sich im Langzeittest das am stärksten vernetzte Kausalmodell – ein Indiz dafür, dass sich ein komplexes Beziehungsgeflecht zwischen den einzelnen Vorstellungsbereichen aufgebaut hat.“ (S. 21) - Komplexes Beziehungsgeflecht „kann als Ursache für den signifikanten Mittelwertanstieg im Langzeittest angesehen werden“ (S. 22), was auch aus den Concept Maps zu schließen ist. - Fragebogen und Concept Map blieben auch in Klasse 8 bestehen – Reliabilitäten blieben bestehen. (S. 21) - Den SuS „der 8. Klasse ohne eine traditionelle Einführung scheint es leichter zu fallen, auf ein makroskopisches Denken in der Mikrowelt zu verzichten“. (S. 23) - Es „gibt eine mit dem Quadrat gekennzeichnete Klasse, deren Probanden einerseits ein angemessenes Modellverständnis besitzen (...) andererseits auf die Übertragung makroskopischer Denkweisen in die Mikrowelt verzichtend argumentieren [die aber] (...) unsicher im mikroskopischen Bereich“ sind. (S. 24) Ähnliches, nur komplementär auch als Dreieck dargestellt - Nach latenter Klassenanalyse wird klar: zunächst haben in Klasse 8 kaum adäquate Metakonzepte bestanden, nach drei Monaten sehr wohl. (S. 24) - Beim Lehrer ohne Metadiskussi- 				
--	--	--	--	--	--	--

			<p>on, wo „die gewünschten Diskussionen und Reflexionen über die Natur der Modellobjekte, über die Natur der Modelle sowie über die Natur des wissenschaftlichen Modellierungsprozesses in der Lerngruppe „2“ nicht stattfanden. Demzufolge ist eine Eigendynamik in dieser Lerngruppe in der Art festzustellen, dass die Schülerinnen und Schüler die Attribute aus der Erfahrungswelt auf die zu untersuchenden Entitäten nach dem Unterricht übertragen. Ein solches Ergebnis lässt vermuten, dass erst durch einen unbedachten Umgang mit den Teilchen im naturwissenschaftlichen Unterricht die Fehlvorstellung von der Übertragung inadäquater makroskopischer Eigenschaften aufgebaut wird.“ Dem steht der metadiskutierende Lehrer gegenüber. Bei diesen SuS stellen sich erwartungskonform eher Metakonzepte ein. (S. 24)</p> <ul style="list-style-type: none"> - „Zusammenfassend lassen die Daten den Schluss zu, dass durch einen Unterricht mit Modellen ein wesentlicher Beitrag zum Verstehen der Teilchenstruktur der Materie geleistet werden kann, indem ein metakonzeptuelles Denken und Argumentieren gefördert wird.“ (S. 24) 				
20	2	Mikelskis-Seifert und Leisner (2004)	<ul style="list-style-type: none"> - „der Übergang zur Mikrowelt bei vielen Schülerinnen und Schülern von makroskopischen Denkweisen dominiert wird.“ (S. 123) - „Eigenschaften der makroskopischen Körper, wie Form und Farbe, werden auf die kleinsten Teilchen übertragen bzw. es wird ihnen eine Temperatur zugeschrieben. - Es besteht die Vorstellung, dass sich zwischen den Teilchen eines Stoffes Luft befindet. - Die Bewegung der Teilchen wird mit Reibung verbunden und soll somit nach gewisser Zeit von allein 	<ul style="list-style-type: none"> - „Kann in einer fächerübergreifenden Intensivwoche, in deren Fokus die Modellierung der Teilchenstruktur stand, ein derartiges Modellbewusstsein angebahnt werden?“ (S. 123) - Ist eine „Anbahnung eines Modellbewusstseins, auf dessen Basis in dem weiteren naturwissenschaftlichen Unterricht eine Modellkompetenz aufgebaut werden kann“, möglich? (S. 123) 	<ul style="list-style-type: none"> - Wirksamkeitsanalyse als Pre-Post-Follow-up-Design (Follow up nach sechs Monaten), Fragebogen (mehrere Skalen zu verschiedenen physikalischen Vorstellungsbereichen und Problemlöseaufgaben + Fragebogen zu Teilchenvorstellungen) und Concept Map und Audioaufnahmen (S. 19 f.), 15 Interviews (Transferfragen), Lerntagebücher (Akzeptanz der Projektwoche) (S. 124), Varianz- und Korrelationsanalysen, Kausalanalysen mithilfe von 	<p><i>N</i> = ca. 120 Schülerinnen und Schüler aus Klasse 8 dargestellt, KEINE Kontrollgruppe!</p>	8. Klasse Gymnasium (Brandenburg)

		<p>aufhören.</p> <ul style="list-style-type: none"> - Zwischen den Teilchen eines Stoffes soll sich derselbe Stoff in kontinuierlicher Form befinden (zum Beispiel Wasser zwischen den Wasserteilchen).“ (S. 123) - „(...) führen diese Fehlvorstellungen zu einem unreflektierten Vermischen von Modell und Realität eines Phänomens sowie zu einer fehlenden Trennschärfe zwischen jeweils erlaubten und unerlaubten bildhaften und sprachlichen Modelldarstellungen.“ (S. 123) - Erster Schritt: „Sensibilisierung ... anhand selbstgebastelter gegenständlicher Modelle – Alltagsmodelle“ (Vorwissen aktivieren) & „Black-Boxen“ (Modellwissen ist hypothetisch) (S. 123) - Zweiter Schritt: „Einführung in die Mikrowelt, in der die Lernenden die Grenzen der direkten Wahrnehmung erfahren sollten“ (Ziel: Trennung von Erfahrungs- und Modellwelt) (S. 124) - Dritter Schritt: Stationenlernen: Arbeitsaufträge zur selbstständigen Erarbeitung von „Diffusion, das Verdunsten, die Kristallisation sowie die Volumenzunahme durch Erwärmen in beiden Welten“ (S. 124) - Vierter Schritt: „Nachdenken über die vorgenommenen Modellierungen“ bzw. Reflexion über „entwickelte Modellvorstellungen“ und Anwendung auf anderes Phänomen (S. 124) - Es gibt schon zum Pretest SuS mit „relativ angemessene[n] Kenntnisse[n] über Kräfte und Abstände zwischen den Teilchen oder über eine Teilchenbewegung. Im Gegensatz zu diesen Denkweisen, die man dem mikroskopischen Bereich zuordnet, sind die Modellvorstellungen unangemessen. Nach der Projektwoche zeigen sich die 		<p>LISREL. LatentClass-Analysen mit WINMIRA sowie Auswertungen auf der Basis von Modalnetzen sowie Schwerpunkt auf Triangulation</p>		
--	--	---	--	--	--	--

			<p>erwünschten positiven Effekte in allen Vorstellungsbereichen, die höchst signifikant sind. Die Lerneffekte im Modelldenken weisen darauf hin, dass sich die naiv-realistische Sichtweise bei den Schülerinnen und Schülern in die Richtung einer hypothetisch-realistischen Position veränderte.“ (S. 124)</p> <ul style="list-style-type: none"> - SuS sind vorsichtiger mit makroskopischen Denkweisen auf submikroskopischer Ebene. (S. 124 f.) - Alle Ergebnisse auch mit Langzeitwirkung (S. 125) 				
21	2	Mikelskis-Seifert et al. (2005)	<ul style="list-style-type: none"> - Trennung von Erfahrungs- und Modellwelt wichtig für SuS (S. 30) - Modellaspekte explizit behandeln, auch hinsichtlich dynamischer Computermodelle (S. 30) - Vollmer (1998) nennt Dinge in anschaulicher Menschengröße „Mesokosmos“. (S. 30) - Größere bzw. kleinere bzw. abstraktere Phänomene benötigen bei der Betrachtung Hilfsmittel. (S. 30) - Für SuS sind „Diskussionen über Erkennbarkeit, Realität und Modellcharakter im Unterricht notwendig“, so Studien von „(Peuckert et al., 2000; Mikelskis-Seifert, 2002)“. (S. 31) - „explizite wissenschafts- und erkenntnistheoretische Reflexion“ wichtig (S. 31) - Ansätze zum Modelldenken: „Schlichting, 1977; Kuhn, 1977; Children Learning in Science Project – CLIS, 1987; Vollebregt, 1997, 1998; u.a.)“ (S. 31) - Empirische Lehr-Lern-Forschung gerade im Bereich der Modelle und Modellbildung, Schwächen im Physikunterricht (z. B. für den submikroskopischen Bereich: Andersson, 1990; Griffiths & Preston, 1992; Treagust et al., 2002, u.a.). (S. 32) 	./.	./.	./.	./.

			<ul style="list-style-type: none"> - „trotz der Bedeutsamkeit der Modelle für das Verstehen von Naturwissenschaften das Modelldenken der Lernenden von Fehlvorstellungen dominiert wird.“ (S. 32) - SuS vermischen Modell- und Realitätsebene oft. (S. 32) - Naiv-realistische erkenntnistheoretische Sichtweisen der Lehrkräfte übertragen sich auf SuS. (S. 33) - Wüsteneck (1963) und Stachowiak (1973): „Erkenntnisgewinnung mit Hilfe von Modellen als dreistellige Subjekt-Modell-Objekt-Relation“ (S. 35), worauf Kircher (1995) aufbaut, ergänzt um Analogien. - Es gibt zwei zentrale Tätigkeiten im Unterricht: <ul style="list-style-type: none"> „1) das Formulieren von Hypothesen, das Erstellen von Prognosen mit dem Modell und das Vergleichen der Prognosen mit der Wirklichkeit sowie 2) das Hinterfragen der Gültigkeit des Modells, das Aufzeigen der Grenzen des Modells und das Reflektieren über die eigene Vorgehensweise.“ (S. 36) - Auch dynamisches Modellieren mit „z. B. Dynasys, Stella, Powersim, ...“ ist wichtig. (S. 36) 				
22	2	Nocke (2001)	<ul style="list-style-type: none"> - Es geht um aktive Gegenstromsysteme über aktiven Transport. (S. 39) - Def. Gegenstromprinzip: wenn „Stoffe in entgegengesetzter Richtung aneinander vorbeiströmen und dabei Substanz oder Wärme von der einen in die andere Flussrichtung übertritt“ (bspw. Harnkonzentrierung oder Temperaturregulation) (S. 39) - Artikel aufgebaut nach Spielregeln 	./.	./.	./.	./.
23	2	Rueckl und Ebinghaus (1989)	<ul style="list-style-type: none"> - Es geht um ein Konzept für die gymnasiale Oberstufe. - „Das Wellenmodell nach De Broglie, sofern es zum Einsatz 	./.	./.	./.	./.

			kommt, erscheint zumeist ohne Bezug zum Bohrschen Modell“, die Hinführung fehlt, daraus entsteht eine „Erklärungslücke“. (S. 21)				
24	2	Schneider (1981)	<ul style="list-style-type: none"> - Richtigkeit der Modellart hängt vom Ziel ab und kann sich ergänzen: Geht es um das anatomische Organ, reicht ein anatomisches Modell, geht es hingegen um die Funktion, bedarf es eines Funktionsmodells. (S. 259) - „Ein Modell soll nicht nur deskriptiv auszuwerten sein, also nicht nur der statischen Demonstration morphologischer Strukturen dienen – zumal Vergrößerung und Farbänderung auch verfälschen –, sondern es soll als Arbeitsmittel einsetzbar sein.“ (S. 259) - Wirbelsäulenreaktionsmodell stark abstrahiert, aber gut für Veranschaulichung der Stoßdämpfung der S-Form (S. 259) - Quantifizierung wenn möglich wichtig in höheren Klassen, auch wenn es sich nur um analoge und nicht reale Messung handelt (S. 259) - „arbeitsunterrichtliche Beschäftigung mit Medien jenen Forderungen, die Vester 1978 für einen sinnvollen und wirksamen Einsatz von Medien fordert: ‚Verdrahten – verankern – vernetzen‘.“ (S. 259) - Es geht dem Autor um die Greifbarmachung der menschlichen Evolution, anstelle eines rein theoretischen, von wissenschaftlichen Namen geprägten Unterrichts (Beispiele: Australopithecus-Schädel und Wirbelsäulenreaktion). (S. 260) 	./.	./.	./.	./.
25	2	Steinbeck und Markert (1989)	<ul style="list-style-type: none"> - „Lernziel (...) Kenntnis von Begriffen wie ‚System‘, ‚Regelung‘ usw.“ ist verschwunden, Kybernetik führt ein Schattendasein, obwohl Regelung noch immer Thema. (S. 39) - Grund: weil „es vielen Biologie- 	./.	./.	./.	./.

			<p>lehren gerade im Mittelstufenunterricht nicht gelingt, den Schülern zu vermitteln, daß sie durch die Verwendung kybernetischer Ansätze zu einem vertieften Verständnis vieler biologischer Sachverhalte gelangen können“. (S. 39)</p> <ul style="list-style-type: none"> - Ziel des Textes: Ursachen für schlechten Erfolg bei Kybernetik werden diskutiert und ein Populationsdichtemodell für Kl. 8 und LKs vorgestellt. (S. 39) - Wie alle Modelle geben Regelkreise die Wirklichkeit nicht vollständig wieder. (S. 39) - Man unterscheide bei Regelkreisen strengstens zwischen: <ul style="list-style-type: none"> „a) Größen: („Größen des Hauptstromes“), Störgröße, Stellgröße und Regelgröße besitzen immer die gleiche Einheit. b) Werte: Die Werte, Istwert und Stellwert, können gewissermaßen mit den Informationen bzgl. der Größen gleichgesetzt werden (...) c) Glieder“ (S. 39) - Ein Problem in Schulbüchern ist, dass die drei zu trennenden Regelkreiselemente weder zeichnerisch noch über eine andere Schriftart voneinander getrennt werden, ein anderes, dass das Bsp. zur Einführung – die Heizung – nichts mit Bio zu tun hat. (S. 39) - Populationswachstum nicht gut durchgeführt - Besser: sich an Begriffsunterscheidung von Schaefer (1972) zu halten (S. 40) - Empfehlung: Beginne bei der Einführung mit dem einfachsten: Größen und Glieder des Hauptstromes. (S. 40) - Analogie Kolbenpöber/Populationswachstum (S. 40): finde ich nicht gut, weil kompliziert. 				
26	2	Streller und Bolte (2008)	- Unterrichtsthema nach Bildungsplänen Luft(-verschmutzung), hier	./.	./.	./.	./.

		<p>Wetter und Klima, verbunden mit den Alltagserfahrungen (emotionalisieren) (S. 11)</p> <ul style="list-style-type: none">- „Ausgehend von einem beeindruckenden Wetterphänomen, dem Auftreten von Tornados und der Zunahme des Auftretens dieser Unwetter, werden in der Sequenz die Entstehung von Tornados erarbeitet und die Zusammensetzung der Atmosphäre sowie die Veränderung dieser durch natürliche und anthropogene Einflüsse thematisiert.“ (S. 11)- SuS „sensibilisieren und ihnen damit Möglichkeiten eröffnen zu erkennen, wie man als aktives Mitglied unserer Gesellschaft Entscheidungsprozesse mitgestalten kann“ (S. 11)-Ziel: mit einfachen Versuchen Informationen, Daten usw. erhalten und damit „sachgerechtes Urteilen“ (S. 12)- IPN-Interessenstudie: großes Interesse, v. a. von Mädchen, an Naturphänomenen (S. 12)- Zusammenhang Wetter Klima und -erwärmung (S. 14)- „Dies entspricht dem prototypischen Ablauf naturwissenschaftlicher Arbeitsweise: Ausgehend von Fragestellungen konnten die Schüler/-innen ihre Ideen äußern, Vermutungen vorbringen und überlegen, welche Informationen sie benötigen, um die Fragen zu beantworten.“ (S. 14)- „(...) in der Auswertungsphase zunächst folgende allgemeine Schlussfolgerungen, die aber jede für sich genommen die Tornadoentstehung noch nicht erklärten:<ol style="list-style-type: none">1. Kühlt sehr feuchte Luft ab, kondensiert der Wasserdampf und der Luftdruck sinkt.2. Warme Luft steigt nach oben.3. Warme Luft hat eine geringere Dichte als kalte Luft und damit				
--	--	--	--	--	--	--

			<p>einen geringeren Luftdruck.</p> <p>4. Der Luftdruck nimmt mit der Höhe ab.</p> <p>5. Wind entsteht durch den Ausgleich von Druckunterschieden." (S. 14)</p> <p>- „Kalte, trockene Luft aus dem Gebirge schiebt sich über warme, feuchte Luft vom Meer; an der Luftmassengrenze kondensiert die Feuchtigkeit, die in der warmen Luft vom Meer enthalten ist: Eine Gewitterwolke entsteht. Die warme Luft steigt auf, gleichzeitig sinkt die kalte Luft ab. Die unterschiedlichen Winde in den verschiedenen Höhen und die sogenannten ‚Scherkräfte‘, die durch die Bewegung der Luftmassen aneinander entstehen, führen dazu, dass eine rotierende Luftsäule entsteht.“</p>				
27	2	Touché (1989)	<p>- Versuch einer Angleichung der physikalischen und chemischen Sichtweise auf Atommodelle</p> <p>- Für die SuS ist so nicht mehr unbedingt zu erkennen, dass es sich um den gleichen Inhalt handelt, dabei wäre es für die SuS wichtig zu erkennen, dass es sich um Beziehungen handelt. (S. 2)</p> <p>- „Das Teilchen sollte dann die folgenden Modelleigenschaften besitzen:</p> <p>a) es besitzt eine Masse (notwendig, um ihm kinetische Energie zuzuordnen zu können)</p> <p>b) es hat Kugelform (nicht notwendig, aber aus Gründen der Einfachheit wünschenswert)</p> <p>c) Teilchen verschiedener Stoffe haben unterschiedliche Durchmesser (nicht notwendig für die Physik, wohl aber für die Chemie)</p> <p>d) es ist beweglich (notwendig, um ihm eine kinetische Energie zuzuordnen zu können)</p> <p>e) es übt anziehende Kräfte aus auf Teilchen desselben oder eines anderen Stoffes (notwendig und</p>	./.	./.	./.	./.

		<p>hinreichend, um Kohäsion und Adhäsion in der Physik und Kräfte im Kristallgitter deuten zu können.“ (S. 2 f.)</p> <p>c) Sowie der Vorschlag, Ethanol + Wasser und Erbsen + Senfkörner können zu trivialen Vorstellungen führen. (S. 3)</p> <ul style="list-style-type: none">- Übergang von Teilchen- zu Atommodell gut dargestellt (Teilchenbegriff wird beim Massenverhältnis „unverzichtbar“, ab „Aggregatzustände“ spräche man im Sinne Daltons besser von Atomen. (S. 3)- „Elemente bestehen aus kleinsten Teilchen. Sie heißen Atome.- Die Atome eines Elements sind untereinander gleich. Sie haben jedoch andere Eigenschaften als die Stoffe, die sich aus ihnen aufbauen. Atome haben keine Stoffeigenschaften wie z. B. Farbe, Glanz, Siede- oder Schmelztemperatur.- Die Eigenschaften der Stoffe beruhen auf der Art der Atome und auf deren Anordnung.- Die Atome üben Kräfte aufeinander aus.- Die Atome bleiben bei chemischen Reaktionen erhalten, sie werden nicht neu geschaffen oder vernichtet.- Verbindungen setzen sich aus den Atomen verschiedener Elemente zusammen.“ (S. 3)- Thomson-Atommodell: „Die Atome stellen kugelförmige, gleichmäßig mit Masse und positiver Ladung erfüllte Gebilde dar. Darin eingebettet liegen die sehr viel kleineren, negativ geladenen Elektronen, gebunden durch elektrostatische Kräfte.“ (S. 4)- „Kern-Hülle-Modell- Atome sind nicht massiv, sondern bestehen aus einem positiv geladenen Kern und einer negativ geladenen Hülle, die die Elektronen enthält.				
--	--	--	--	--	--	--

			<ul style="list-style-type: none"> - Der für α-Teilchen undurchdringliche Kern hat einen Durchmesser von 10^{-14}m, die Hülle hat einen solchen von 10^9m. - Atomkerne enthalten Protonen ($m = 1u$, $e = 1,6 \cdot 10^{-19}$ As) und Neutronen ($m = 1u$, $e = 0$ As). - Die Anzahl der Protonen im Kern und die der Elektronen in der Hülle ist gleich. <p>Energetische Deutung im Energiestufenmodell</p> <ul style="list-style-type: none"> - Elektronen innerhalb eines Atoms besitzen bestimmte Energien. - Energetisch etwa gleichwertige Elektronen gehören zu einer Energiestufe. <p>Räumliche Deutung im Schalenmodell</p> <ul style="list-style-type: none"> - Elektronen halten sich gruppenweise in bestimmten Entfernungen vom Kern auf. - Elektronen, die sich etwa in der gleichen Entfernung vom Kern aufhalten, gehören zu einer Schale.“ (S. 4 f.) 				
28	2	Winnenburg (2000)	<ul style="list-style-type: none"> - „Verstehen‘ ist dabei eine besondere Art von Erkenntnis, eine sokratische, die die Gründe für das Erkannte weiß.“ (S. 324) - „Lehrende in der Schule und andernorts werden heutzutage häufig mit einem von den Medien erzeugten Umfeld von Halbwissen konfrontiert. Es ist daher zu einer wichtigen didaktischen Aufgabe von Unterricht geworden, den Schülern Medienkompetenz im Sinne von Fähigkeit zur kritischen Nutzung und Bewertung von Presse, Funk, Fernsehen, Internet, etc. zu vermitteln.“ (S. 324) - „Zum unterrichtlichen Vorgehen gehört dabei auch das Anzweifeln und Infragestellen der empfundenen oder entdeckten Erscheinungen.“ Dies führt zu kognitiven Konflikten, die in einen Gleichgewichtszu- 	./.	./.	./.	./.

			<p>stand gebracht werden wollen. (S. 325)</p> <p>- „Zur tatsächlichen Einsicht ist es notwendig, einen Standortwechsel nach außerhalb des Sonnensystems zu vollziehen“, obwohl man meist von sich ausgeht. (S. 325)</p>				
29	2	Cohen und Yarden (2009)	<p>- Die Lehrkräfte änderten den “way of teaching the cell topic” nur oberflächlich: “they did not undergo any deep change” (131) hin zu einer langfristigen Vermittlung von Zellen, verbunden auch mit anderen Themen.</p> <p>- “The teachers were also found to have no PCK capacity to integrate biological phenomena at the macro-level with their cellular explanations.” (131)</p> <p>- “a duality was identified among the teachers with regards to relating macro- and micro-levels in biology and in chemistry.” (131)</p> <p>- “comprehension difficulties with regards to the living cell” (132) wurden in Israel und der Welt (Flores et al. 2003; Lewis et al. 2000; Verhoeff 2003 (evtl Präkonzepte) festgestellt.</p> <p>- Der Ansatz hat das Potenzial, “providing an opportunity to form meaningful relationships between the cellular level and other organizational levels”. (S. 132)</p> <p>- Dieser komplexe neue Ansatz “requires the development of appropriate teachers’ PCK”. (S. 132)</p> <p>- “Teaching the Living Cell Topic as Part of the Living World” (S. 132)</p> <p>- Lehrkräfte untersuchten eigentlich hauptsächlich 7- und 8.-Klässler Chemie</p> <p>- In den Interviews wird das Thema immer wieder als langweilig beschrieben. (S. 141)</p>	<p>Evaluation nach Curriculumānderung (langfristige Vermittlung von Zellen verbunden auch mit anderen Themen): PCK zu Zellen (131)</p> <p>Ziel: „The purpose of this study was to gain a better understanding of the junior-high-school science and technology teachers’ PCK with regards to the living cell topic and with regards to the formation of macro-micro relations.“ (S. 132)</p>	<p>1) “Teachers’ responses to authentic students’ explanations of biological phenomena” (137) orientiert an den PCK-Dimensionen von Magnusson et al. (1999). Die Schülerantworten mussten gescort werden. (S. 137) Punktwert im Vgl. zu wessen Meinung ergibt PCK (S. 137).</p> <p>2) Semistrukturierte Interviews ($N = 6$) (S. 138).</p> <p>3) Zeichnungen zum Curriculaufbau einschätzen (S. 139)</p> <p>4) Analyse von Klassenarbeiten der Lehrkräfte</p> <p>5) Fragebogen zu Erwartungen an den Workshop (S. 139)</p> <p>6) Raten von nicht bekannten Fragen für einen Schultest ((nicht)geeignet für Klasse 7-9)</p>	<p>„focus groups ($n = 59$)“, “workshop ($n = 12$)”, $n = 6$ erfahrene Lehrkräfte wurden interviewt (131)</p>	<p>Erfahrene Naturwissenschafts- und Technik-Junior-Highschool-Lehrer</p>

			<ul style="list-style-type: none"> - Der neue Lehransatz war den Lehrkräften nur halb verständlich (S. 142), aber manche verstanden auch, dass es sich um einen „spiral topic“ handele. (142) - Ein riesiger Mangel an adäquatem Lernmaterial wurde festgestellt. (S. 144) - Folgende Präkonzepte waren Probanden bekannt: <ul style="list-style-type: none"> - Visualisierungsprobleme hinsichtlich des mikroskopischen Bildes (S. 145) - Zellen nicht mit den Sinnen erfahrbar, weil zu klein (S. 145) - 3D-Problem (S. 145) - Schwierigkeiten mit dem Verständnis von zellulären Prozessen (S. 145) - Präkonzept: Zellen haben kleine Lungen. (S. 145: Dreyfus and Jungwirth (1988)) - Entsprechen Präkonzepten von „Dreyfus and Jungwirth 1988, 1989; Flores et al. 2003“ (S. 145)). - „Teachers’ Duality with Regard to Relating Macro- and Micro-Levels in Biology and in Chemistry” (S. 148): Chemie und Bio werden in Israel immer zusammen unterrichtet. - “The question is why the teachers naturally connect the macro- and micro-levels in chemistry and do not tend to do so in biology.” (S. 148 f.): Sie waren sich über die Verbindung zu (Dis-)Kontinuum bzw. Mikro-/Makro in Chemie nicht bewusst. (S. 149) - Diskussion, ob Magnussons 5 PCK-Facetten ein Ganzes oder Teile getrennt sind (S. 150) - Tiefgreifende Änderungen (CC) blieben trotz der Curriculumänderung aus. (150) 				
30	2	Crawford und Cullin (2004)	- “(a) prospective teachers became more articulate with the language of modelling; and	(1) What do prospective teachers understand about models and modelling in	Vor- und Nachmodul, in depth-Interviews, Fragebogen: open-ended questions nach Gross-	N = 14 künftige (prospective) Lehrer, Tiefeninterviews mit N = 6	Künftige Lehrkräfte/Naturwissenschafts-Lehramtsstudierende

		<p>(b) the module enabled prospective teachers to think critically about aspects of modelling.“ (S. 1379)</p> <ul style="list-style-type: none"> - “used qualitative methods to explore the influence of this module on these novice teachers’ understandings of the nature and use of scientific models.” (S. 1380) - “authors are particularly concerned about prospective teachers’ understandings and use of scientific models related to the nature of science and scientific inquiry.” (S. 1380) - Justi und Gilbert (2001) “most teachers do not possess comprehensive knowledge and skills related to building models required for supporting students in learning science.” - “The role of models and modelling in science represents an important and often neglected aspect of scientific inquiry and the nature of science (American Association for the Advancement of Science [AAAS] 1993, Gilbert 1995, Hodson 1993).” (S. 1381) - “Models are integral to thinking and working scientifically because models are ‘science’s products, methods, and its major learning and teaching tools’ (Gilbert 1993: 9–10).” (S. 1381) - Modelle historisch und philosophisch: „Black 1962, Giere 1990, Hesse 1966“ (S. 1381) - „Precollege students may understand the empirical aspect of scientific claims, but they confuse scientists’ use of new perspectives in looking at data, instead of simply changing models and theories as new evidence comes to light (Bell et al. 2003).” (S. 1381) - “the iterative aspect of modelling connects to understanding the nature of science.” (S. 1381) - Es geht um “simulation models”. 	<p>science?</p> <p>(2) How do their understandings change, following building and testing dynamic computer models? and</p> <p>(3) What are prospective teachers’ intentions to teach about scientific models? (S. 1379)</p> <p>Alles hinsichtlich dynamischer Modelle, konstruiert in Model-IT</p>	<p>light (1991), bis auf 7/8,:</p> <p>“1. What is a scientific model? 2. What is the purpose of a scientific model? 3. When making a model, what do you have to keep in mind or think about? 4. How close does a model have to be to the thing itself? 5. Would a scientist ever change a model? If so, why? If not, why not? 6. Can a scientist have more than one model for the same thing? If so, why? If not, why not? 7. Is teaching about models important in your area of science? Why or why not? 8. Do you intend to teach students about models and modelling? Why or why not?” (S. 1390)</p> <p>Schließlich Reflektion über Modellerfahrung als “journal responses“, auch verbunden mit diesen Fragen:</p> <p>“Include your thoughts on:</p> <ol style="list-style-type: none"> 1. the importance of involving your students in inquiry; 2. the importance of involving your students in modelling; 3. your level of comfort in designing activities for students in which they would engage in inquiry and modelling; 4. difficulties you perceive in engaging your own students in inquiry and modelling.” (S. 1391) 		
--	--	--	--	---	--	--

			<p>(S. 1382)</p> <ul style="list-style-type: none"> - Begründung für CFA-Veränderungen: “scientists continuously revise models, as they probe new phenomena, collect additional data, and interpret the same data in new ways.” (S. 1382) - Gute Zusammenfassung zu Überzeugungen von Schülern und Lehrkräften zu Modellen ab S. 1382 - Modellieren von TerrAqua bzw. <i>Bottle Biology</i> aus mit Model-IT (S. 1388) - „The test mode of <i>Model-It</i> allowed prospective teachers to think critically about effects of the variables on their system in the real world.” (S. 1388) - Es gab Fortschritt, aber diese schlugen sich nicht im Kategoriensystem von Grosslight (1991) nieder. (S. 1399) 				
31	2	Drechsler und van Driel (2008)	<ul style="list-style-type: none"> - Es gab Lehrkräfte “categorised as student and model oriented” und “teacher oriented and micro/macro-level oriented”. (S. 625) Zweitere argumentierten „a macroscopic or phenomenological level, and a microscopic or particulate level”. - “Knowledge of models and their use, and recognition of their limitations, would allow students to gain a better understanding both of scientific knowledge and of the nature of science; that is, how scientific knowledge is achieved.” (S. 612; könnte man zitieren) - „three distinct models can be identified. These models are (a) the ‘ancient’ model, (b) the Arrhenius model, and (c) the Brønsted model.” (S. 613) - “According to van Driel et al. (1998), two key elements of PCK are essential in all research about teachers’ knowledge. These elements are (1) teachers’ knowledge about specific conceptions and learning difficulties with respect to 	<ul style="list-style-type: none"> - “1. What is the content of experienced teachers’ PCK about students’ difficulties in understanding acids and bases? 2. What is the content of experienced teachers’ PCK of teaching strategies they consider useful to help students overcome such difficulties; in particular, do they intend to use models of acids and bases in their teaching? 3. How did the teachers perceive (wahrnehmen) their PCK of teaching acids and bases develop until now?” (S. 616) 	<p>Qualitativ: „semi-structured interview” (S. 617):</p> <ol style="list-style-type: none"> 1) Was wurde vom Kurs erinnert? 2) Unterrichtsplanungsänderungen über Jahre hinweg zu Säuren und Basen 3) Diskussion über ihr Schulbuch und Reaktion auf Ausschnitte aus Büchern 4) Schülervorstellungen zu Säuren und Basen wurden diskutiert – Reaktion auf authentische Schüleräußerungen (kennen sie solche Äußerungen, wie darauf reagieren, woher kommen solche Vorstellungen, glauben Sie, die Schüler verstehen, warum verschiedene Modelle bei Säuren und Basen eingesetzt werden und ob Schüler diese Unterschiede kennen? Zeichnen story-lines zur Zufriedenheitsentwicklung hinsichtlich Vermittlung von Säuren und Basen) 	N = 9 erfahrene Chemielehrkräfte	erfahrene Chemielehrkräfte

			<p>particular content and (2) teachers' knowledge about representations and teaching strategies." (S. 615)</p> <ul style="list-style-type: none"> - Magnusson et al. (1999) zu PCK (vgl. Crawford und Cullin (2004): “(a) orientation toward science teaching, (b) knowledge of science curriculum, (c) knowledge of science assessment, (d) knowledge of students' understanding, and (e) knowledge of instructional strategies” (S. 615) 			
32	2	Eilks et al. (2009a)	<ul style="list-style-type: none"> - „discusses what might occur while working with animations taken from the Internet and how these multimedia illustrations can potentially interact to reinforce rather than resolve students' misconceptions about chemical principles.” (S. 49) - “We know from constructivist theories of learning (e. g., Bodner, 1986) that visual stimuli are not simply captured in learners' heads. They are filtered and then re-interpreted using the framework of the viewer's preconceptions. Students do not anchor into their minds exactly what they see in an easy analogy.” (S. 51) - “Chemistry necessarily tries to explain macroscopic properties and phenomena on the sub-microscopic level.” (S. 51) - “the inaccessibility of the sub-microscopic level to direct human senses also is one of the main problems (e.g. Johnstone, 1991).” (S. 51) - “largely dependent on human imagination to depict and describe.”(S. 51) - Konstruktivistische Lerntheorie: Bodner (1986) - Präkonzept: “Classroom experience frequently reports that the ions 	<ul style="list-style-type: none"> - “The question should be allowed, as to whether all animations used for illustrating scientific concepts are genuinely useful in supporting learning.” (S. 50) - “must reflect on whether any specific illustrations taken from the Internet actually provide us with visual stimuli that are feasible to support learning in the direction of scientific correctness and the intended learning objectives.” (S. 50 f.) 		

		<p>within an electrochemical cell are thought to be 'produced by' the cell by many students. They also state that such particles 'disappear' after being uncharged at the corresponding electrode." (S. 52)</p> <ul style="list-style-type: none">- Präkonzept: "possible gain (or loss) of mass by the two oppositely-charged electrodes. The students only take into account the role of the ions with respect to the macroscopic phenomenon of achieving a flow of electricity. But the students neglect the sub-microscopic character of the ions as particles which must be conserved." (S. 52)- "The loss in mass of the zinc anode and the increase in mass of the copper cathode can be specifically measured. The result can be used to provoke cognitive dissonance and a discussion in those students who neglect or misinterpret the change in mass occurring to the electrodes." (S. 52)- "Johnstone (1991) in his very famous contribution "Why is science difficult to learn?" explained that learning and thinking in chemistry always takes place in an area of conflict between the macroscopic, sub-microscopic, and symbolic level." (S. 52)- Pro Animation: "In principle, animations and simulations can be considered as powerful tools for teaching and learning chemistry. They can foster students' understanding of three-dimensional structures (Williamson & Abraham, 1995) and even help developing learners' spatial ability (Barnea & Dori, 1999). They provide a resource which may help to reduce students' misconceptions about basic chemical principles (Kozma & Russel, 2005b; Sanger & Greenbowe, 2000; Yang, Greenbowe & Andre, 2004) and the use of anima-				
--	--	--	--	--	--	--

		<p>tions has been shown to motivate students' learning of chemistry (Tsui & Treagust, 2004)." (S. 53 f.)</p> <p>- Contra Animation:</p> <p>"1) limited learner attention spans when viewing animations, due to the high transfer rate of information (Ploetzner, Bodemer, & Neudert, 2008),</p> <p>2) inadequate meta-cognitive competencies (especially in non-linear learning environments) (Schwartz, Andersen, Hong, Howard, & McGee, 2004; Azevedo, 2004),</p> <p>3) a lack of prior knowledge (Shapiro, 1999) or</p> <p>4) too little ability in recognizing or using proper spatial relations (Lee, 2007; Huk, 2007)." (S. 54)</p> <p>- "Schnotz and Bannert (2003) discussed the fact that picture-use in multimedia learning processes are not necessarily beneficial to learning in every case." (S. 54)</p> <p>- "Therefore, if effective learning is to take place, illustrations and visual aids must be structured to take the learner's preknowledge of a given topic into account. Additionally, the relationship towards the scientifically accepted explanation and – if relevant – the model nature of the chosen explanation (Kelly, 2005, as discussed in Falvo, 2008) must also be taken into account." (S. 54)</p> <p>- Anweisung an Studierende: "Watch the animation and write down what you see!" (S. 55)</p> <p>- "But why then do we as experienced chemists 'see' what we believe we see? Our 'expert knowledge' leads us to perceive what we want to see: an animated illustration of the Daniell voltaic cell. Within seconds, we reconstruct our knowledge using the impulse of the animated picture in order to</p>			
--	--	---	--	--	--

			<p>obtain the 'correct' view. The observed content is no longer of interest to us; unfortunately we can't expect the same for our inexperienced students." (S. 57)</p> <p>- Die beschriebenen Probleme wären dann kein Problem, "if a model is sufficiently discussed and reflected upon in the classroom, there may no longer be any misunderstandings". (S. 58)</p> <p>- "In most cases, students do not have a sufficiently-developed understanding of scientific models and modelling (Grosslight, Unger, Jay, & Smith, 1991)." (S. 58) Und Lehrer auch nicht (Van Driel & Verloop, 1999; Harrison, 2000; Justi & Gilbert, 2002a and b; Sprotte & Eilks, 2007)</p> <p>- Präkonzeptaufzählung S. 58</p> <p>- „which kind of simplification is acceptable, and which will only serve to nurture students' alternative beliefs?" (S. 58 f.)</p> <p>- Entstehung TPCK: "Mishra & Koehler (2006) suggested picking up technology as a fourth, interacting dimension in the discussion of PCK. They suggested the construct of <i>Technological Pedagogical Content Knowledge</i> (TPCK)." (S. 62)</p>			
33	2	Garritz (2013)	<p>- Es geht um "awareness of the 'historicity'" (S. 1787)</p> <p>- "The establishment of a theory is accompanied with philosophical interpretations all the way" (S. 1787), was auch auf einen spezifischen Umgang mit Modellen im Unterricht verweist.</p> <p>- „it gives excellent results in teaching and learning to bring to the foreground the complexity that surrounds the development of ideas in science..." (S. 1787)</p> <p>- "created to describe an abstract atomic world far away from daily experience, its actual impact on our</p>	Nicht vorhanden, evtl. ist Historisierung und Philosophisierung von Quantenmechanik im Unterricht förderlich oder nicht?		

		<p>lives could hardly be greater, because of its impact on modern technology-lasers, transistors and semiconductors.” (S. 1788)</p> <p>- “Without quantum mechanics there would be no global economy to speak of, because the electronics revolution that brought us to the computer age is a child of quantum mechanics.” (S. 1788)</p> <p>- Die Frage, ob historische Modelle wie Bohr geeignet sind, wird aufgeworfen. (S. 1788)</p> <p>- Zählt zunächst Gegner der Historisierung auf (argumentieren mit entstehenden Lernschwierigkeiten, z. B. Fischler and Lichtfeld 1992). (S. 1788)</p> <p>- Historisierung nach Conant (1949) „is a very good one in fostering learning that scientific development is not based on accumulation of data or constitutes a simple linear process”. (S. 1788): Es geht ihm also eher um NoS usw. als um reines Faktenlernen.</p> <p>- „The author of this article thinks that arguments on quantum physics are grounded in premises that are not evidently true, that are complex and counter-intuitive, they are dialectical, although also have a rhetorical character.” (S. 1801)</p> <p>Zitat für Diskontinuum nutzen, obwohl es um Quantenmechanik geht!</p> <p>- <i>Präkonzeptliteratur:</i></p> <p>- Coll, R. K. & Treagust, D. F. (2003a). Learners’ mental models of metallic bonding: A cross-age study. <i>Science & Education</i>, 87(5), 685–707.</p> <p>- Coll, R. K., & Treagust, D. F. (2003b). Investigation of secondary school, undergraduate, and graduate learners’ mental models of ionic bonding. <i>Journal of Research in Science Teaching</i>, 40(5), 464–486.</p> <p>- Justi, R. & Gilbert, J. (2000).</p>				
--	--	---	--	--	--	--

			<p>History and philosophy of science through models: Some challenges in the case of 'the atom'. International Journal of Science Education, 22(9), 993–1009.</p> <p>- Kalkanis, G.; Hadzidaki, P., & Stavrou, D. (2003). An instructional model for a radical conceptual change towards quantum mechanics concepts. Science & Education, 87, 257–280.</p> <p>- Park, E. J. & Light, G. (2009). Identifying atomic structure as a threshold concept: Student mental models and troublesomeness. International Journal of Science Education, 31(2), 895–930.</p> <p>- Zeidler, P. (2000). The epistemological status of theoretical models of molecular structure. International Journal on the Philosophy of Chemistry, 1(1), 17–34.</p> <p>- Bindungspräkonzepte-Untersuchungen aufgezählt auf S. 1790</p>				
34	2	Harrison und Jong (2005)	<p>- Fachliche Themen zu Analogien waren Reaktionsraten, umkehrbare Reaktionen und chemisches Gleichgewicht (S. „1“)</p> <p>- Zitat für Modelle sind Teil von PCK: „PCK components are: (a) knowledge of specific instructional strategies, knowledge of multiple representations (models and analogies) and activities (e.g. demonstrations); and (b) knowledge of student prior knowledge of a topic and known student learning difficulties (see Gess-Newsome & Lederman, 1999).” (S. „1“)</p> <p>- 7 Analogien:</p> <p>- „“busy-highway” analogy and the “dance-commitment room” analogies to illustrate the dynamic nature of reversible reactions and the equilibrium that can exist between reactants and products.” (S. „2“)</p> <p>- “better understand the submicroscopic world by helping the stu-</p>	<p>- “1. What correspondences and differences between a teacher’s intentions and his/her practice can be identified when teaching with analogies?</p> <p>2. How can teacher reflections on their intentions and teaching with analogies enhance their explanations in science?” (S. “1”)</p>	Interview jeweils vor der Stunde zu seinen Intentionen und nach der Stunde zu seiner Reflektion, Fokus: meistens auf Analogien zur Illustration und der Erklärung von chemischen Konzepten (S. „1“)	N = 1 erfahrener Chemielehrer	Erfahrene Chemielehrkräfte

			<p>dents visualize particle interactions.” (S. “2”)</p> <ul style="list-style-type: none"> - “attention to limitations of the analogies by inviting the students to find where the analogies break down.” (S. “2”) -> Modellmethode - “check students’ understanding of the analogies by asking them to retell the links between a particular analog and target in their own words.” (S. “2”) - “aware of students’ difficulties in understanding the “busy-highway” analogy” (S. “2”) - “These results underline the need to pay attention to specific aspects of teaching with analogies in the context of pre-service and in-service teacher education.” 				
35	2	Jong und van Driel (2004)	<ul style="list-style-type: none"> - SuS haben in Chemie Probleme, Themen hinsichtlich makroskopischer, mikroskopischer und symbolischer Bedeutung auseinanderzuhalten. (S. 477), vgl. Johnstone (1991) (S. 479) - “As a consequence, student teachers should be prepared carefully to learn how to teach this difficult issue.” (S. 477) - Ergebnisse weisen darauf hin, dass es durch das Lehren eine Entwicklung beim Wissen um Lehrschwierigkeiten hinsichtlich z. B. zu schneller Schlussfolgerung von Mikro- auf Makrobedeutung gibt, außerdem ein Anstieg von Wissen um Lernschwierigkeiten bei SuS. (S. 477) - Für Lehrkräfte “a formula as H₂O refers to the substance of water as well as to a single molecule of water. Their mental switching between macro-aspects and micro-aspects of science curriculum topics is conducted easily and almost automatically”. (S. 480) - Für SuS “the conceptual demands of shifting between the three meaning domains can be overwhelming”. 	<p>“what development of the student teachers’ PCK can be identified concerning the issue of the macromicro-symbolic meanings of chemistry topics?” (S. 481)</p>	Leitfragengestütztes Prä-post-Unterrichtsgespräch	N = 8	Chemielehramtsstudierende mit Fachchemiemaster

		<p>(S.480)</p> <ul style="list-style-type: none"> - Präkonzepte Teilcheninteraktion: “(e.g., Benson, Wittrock & Bauer, 1993; Lee, Eichinger, Anderson, Berkheimer & Blakeslee, 1993; Treagust, Chittleborough & Mamiya, 2003)” (S. 480) - Präkonzepte zur algebraischen Überinterpretation statt Mikro-makro-Beziehung: (Ben-Zvi, Eylon & Silberstein, 1988; Friedel & Maloney, 1992). (S. 480) - „The latter approach assumes that student teachers learn in a mainly passive way how to teach, whereas learning <i>from</i> teaching means that student teachers learn in an active way, involving real practice situations, to make their learning more meaningful to themselves (cf. Lampert & Loewenberg, 1998).“ (S. 481) - Zum Startinterview wurden die Studierenden nach ihren Planungen gefragt. - Bei den post-unterrichts-Interviews galt das Augenmerk den Lehr- und Lernschwierigkeiten. (S. 482) - Lehrschwierigkeiten umfassen vier Bereiche: <ol style="list-style-type: none"> 1) Zu schnelles Schlussfolgern von der Mikro- zur Makrobedeutung (S. 483) 2) Hauptsächlich Orientierung auf Mikrobedeutung (S. 485) 3) Vermischung von Mikro-Bedeutung und Makrobedeutung (S. 485) 4) Umgang mit Vorschriften zur Notierung von symbolischen Repräsentationen (S. 486) - Wissen über Schülerlernschwierigkeiten (S. 486) <ol style="list-style-type: none"> 1) Falsches Verständnis der Bedeutung von Formeln 2) Falsches Verständnis von Reaktionsgleichungen (S. 487) - “After teaching, much more 			
--	--	---	--	--	--

			<p>difficulties were reported.” (S. 488)</p> <p>– “database of students’ conceptual and reasoning difficulties (CARD) can be found on the website: www.card.unp.ac.za. (S. 490)</p> <p>- “In sum, it is important that student teachers get the opportunity to link their authentic teaching experiences with analyses of relevant articles, for instance, about students’ preconceptions and ways of reasoning in a specific domain (see also Geddis, 1993).” (S. 491)</p>				
36	2	Jong et al. (2005b)	<p>- Es geht um „relationship between phenomena (e.g., properties of substances, physical and chemical processes) and corpuscular entities (e.g., atoms, molecules, ions).” (S. 947)</p> <p>- “After teaching, all preservice teachers demonstrated a deeper understanding of their students’ problems with the use of particle models.” (S. 947)</p> <p>- PCK ist “distinct from general knowledge of pedagogy, educational purposes, and learner characteristics”. (S. 948)</p> <p>- “However, several investigators have pointed out that it is not always possible to make a sharp distinction between PCK and subject matter knowledge (Marks, 1990; Tobin, Tippins & Gallard, 1994).” (S. 948)</p> <p>- “knowledge of science learners concerns student learning of a specific topic and comprises knowledge of students’ learning difficulties, whereas knowledge of instructional strategies includes knowledge of representations (e.g., models, metaphors) and activities (e.g., explications, experiments) for teaching a specific topic.” (S. 948 f.)</p> <p>- “These investigators concluded that a teacher education program can never completely address all</p>	<p>- “How does preservice teachers’ PCK of the use of particle models develop in the context of a specific course module within a chemistry teacher education program?”</p> <p>- “1. What is the content of preservice teachers’ initial PCK of learning difficulties concerning the use of particle models?</p> <p>2. What is the content of preservice teachers’ initial PCK of instructional strategies they consider useful to overcome such learning difficulties?</p> <p>3. What is the content of preservice teachers’ PCK after participating in the course module, including teaching a series of lessons about the use of particle models to help secondary school students understand the relationship between phenomena and corpuscular entities?” (S. 951)</p>	<p>- “The data collected consisted of: (a) the written answers of each individual preservice teacher to the questions and assignments included in the four parts of the module; (b) the reflective lesson reports written by the PTs at the end of the module; and (c) the audiotape recordings of all plenary discussions that took place during the institutional workshops.” (S. 952)</p>	N = 12	preservice chemistry teachers: 1-year postgraduate program

			<p>the components of PCK a teacher needs.” (S. 949)</p> <p>- “To enhance the impact of classroom teaching experiences, preservice teachers should be stimulated to reflect on their own teaching (Osborne, 1998).” (S. 949)</p> <p>- Justi and Gilbert (2002, S. 57): “overview of the specific role of models and modeling in chemical education, ‘there have been very few initiatives to promote the development of teachers’ pedagogical content knowledge in this area’.” (S. 951)</p> <p>- “An experimental course module was developed, which aimed at learning to use particle models to help students in secondary education understand the relationship between observable phenomena and corpuscular entities.” (S. 951)</p> <p>- “This design can be categorized within the group of ‘interactive models’ of teachers’ professional development (Sprinthal, Reiman & Thies-Sprinthal, 1996), and requires strong linkages between institutional activities and classroom practice.” (S. 952)</p> <p>- “Interconnected Model of Teacher Professional Growth (IMTPG; Clarke & Hollingsworth, 2002).“ (S. 961)</p>				
37	2	Justi und van Driel (2005a)	<p>- „action research was included as a strategy to contribute to the development of beginning science teachers’ knowledge.” (S. 198)</p> <p>- Eine postgraduierte Lehrkraft wird untersucht.</p> <p>- Wichtigkeit von Modellen: “(a) models play essential roles in science, fulfilling a series of purposes, such as making abstract entities visible (Francoeur, 1997), and providing a basis for the formulation of predictions (Erduran, 2001; Hesse, 1966; Mainzer, 1999), (b) models and modelling are</p>	<p>“How did the knowledge of this beginning science teacher about models and modelling develop when she participated in a project especially designed with this aim?” (S. 199)</p> <p>Ziel: “to contribute to a better understanding of the development of the personal practical knowledge (Beijaard et al., 2000) of beginning teachers, which may be</p>	<p>- Fragebogen VOMM C (schriftliche Auswertung) und Interviews, geschriebenes Material, Transkripte von Gruppendiskussionen, Forschungsplan und -aufzeichnung der Lehrerin samt Video (S. 203)</p>	N = 1 (Julie)	Zukünftige Chemielehrkräfte

			<p>particularly relevant with respect to all three purposes of science education identified by Hodson (1992), that is, <i>learn science</i>, <i>learn about science</i> and <i>learn to do science</i>, and (c) recent studies have revealed that the knowledge base of science teachers, both experienced and beginning, in this area is often limited or even problematic (Crawford & Cullin, 2002; Harrison, 2001; Justi & Gilbert, 2002a, 2002b, 2003a; van Driel & Verloop, 1999).“ (S. 199 f.)</p> <p>- „Some of the most common teaching models are drawings, simulations, analogies and concrete models.” (S. 200)</p>	<p>useful to design effective teacher education activities.” (S. 199)</p>			
38	2	Justi und van Driel (2005c)	<p>- “science teachers’ content knowledge, curricular knowledge, and pedagogical content knowledge about models and modelling are often incomplete or inadequate.” (S. 549)</p> <p>- “Teachers’ knowledge is of pivotal importance in the design and conduct of teaching situations that may help students to learn science.” (S. 549) -> Es geht um mehr als nur Wissen, sondern eher um das Handeln. Der Zusammenhang ist unklar...</p> <p>- Wissen über die Natur der Naturwissenschaft hinsichtlich Modellen fehlt (Justi and Gilbert, 2003). (S. 550)</p> <p>- In der Lehrkräfteausbildung sollten Lernsituationen, ihr Vorwissen und ihre Erfahrungen mit berücksichtigt werden. (S. 550)</p>	<p>“1. How can the development of teachers’ content knowledge, curricular knowledge, and PCK on models and modelling be fostered?</p> <p>2. How do specific aspects of teachers’ content knowledge, curricular knowledge, and PCK on models and modelling change when they participate in the project?</p> <p>3. How do teachers’ knowledge changes manifest themselves in their classes?</p> <p>4. Which aspects should be taken into account in promoting the development of teachers’ content knowledge, curricular knowledge, and PCK on models and modelling?” (S. 551)</p>	<p>- Interview - Fragebogen VOMM C (Justi und Gilbert 2003)</p>	<p>N = 5 postgraduierte Masterstudierende Chemie/Physik</p>	<p>Lehramtsstudierende</p>
39	2	Justi und van Driel (2005b)	<p>- Es gibt lt. Sprinthal, Reiman und Thies-Sprinthal (1996) drei generelle Erklärungsmodelle für die Entwicklung von Lehrkräften: “the craft, the expert and the interactive model”. (S. 197): the craft: Leh-</p>	<p>- “1. How does the ‘external domain’ influence the teachers’ initial personal domain and their domain of practice?</p> <p>2. How do specific aspects</p>	<p>-Interview -Fragebogen VOMM C (Justi und Gilbert 2003) - Interview</p>	<p>N = 5 postgraduierte Masterstudierende Chemie/Physik</p>	<p>Lehramtsstudierende</p>

			<p>erfahrung: expert model: Lehrkräfte werden von Experten gelehrt, was und wie gelehrt wird. Interactive model: "the teacher as an active participant in the learning process" (Sprinthall, Reiman and Thies Sprinthall (1996, S. 687) als "strong linkages between institutional activities and classroom practice".</p>	<p>of teachers' content knowledge, curricular knowledge and PCK on models and modelling (i.e., their personal domain) change when they participate in the project, due to interactions with the other domains in the IMTPG?" (S. 169)</p>			
40	2	Khourey-Bowers und Fenk (2009)	<p>- "(1) constructivist chemistry professional development (PD) meets the needs of teachers of varying CK, and (2) instruction should connect representational models with alternative conceptions, integrating radical and social constructivism." (S. 437)</p> <p>- Radikaler und sozialer Konstruktivismus als Rahmen</p> <p>- Präkonzepte "The extensive findings of Driver and her colleagues (Driver et al. 1985; Driver et al. 1994) provide a catalog of children's ideas about fundamental concepts in chemistry, including the nature of matter, energy, change of state, and chemical reactions."</p> <p>- Wie man CC hervorrufen kann: "Inclusion of conceptual change strategies, such as explicit discussion of alternative conceptions and implementation of classroom-appropriate activities to address students' conceptions (Harvard-Smithsonian Center for Astrophysics 1995) may encourage teachers to use innovative and active instructional strategies." (S. 442)</p> <p>- "Parallel to the goal of changing alternative conceptions is the goal of changing mental models into analogical models consistent with scientific ways of thinking. Mental (nonscientific) models are personal views of reality, theories, and processes." (S. 442)</p>	<p>- "1) If elementary and middle/secondary grades teachers participate in constructivist PD, which provides instruction in fundamental chemistry concepts, there will be no difference in the CK of either group.</p> <p>2. If elementary and middle/secondary grades teachers participate in constructivist PD which explicitly provides instruction in scientific models, there will be no difference in the use or understanding of multiple levels of scientific models of either group.</p> <p>3. If elementary and middle/secondary grades teachers participate in constructivist PD, neither group will show gains in PSTE.</p> <p>4. If elementary and middle/secondary grades teachers participate in constructivist PD, neither group will demonstrate enhanced PCK through classroom practices which (a) elicit alternative conceptions, (b) provide for public exposition of alternative conceptions, (c) use a bridging activity to help students develop more scientifically consistent ways of thinking, and (d)</p>	<p>- t-Test (S. 445)</p> <p>- Befragung nach CC-Momenten im Unterricht (Selbstbericht)</p>	<p>N = 69 Grundschul- und Sekundarstufe I-Lehrkräfte</p>	<p>Chemie-Lehrkräfte</p>

				use analogical models to convey scientific concepts.” (S. 439)			
41	2	Liang et al. (2011)	<p>- „we found that students changed their models about gas behavior when the orientation of the apparatus changed, and that they had difficulty using a consistent model to answer all of the questions, especially those questions involving a change of pressure.” (S. 238)</p> <p>- Präkonzepte: “molecules are not in constant motion, the particles in a gas are not uniformly distributed, molecules move in gases and liquids but not in solids, and so on (Novick and Nussbaum, 1981; Lee et al., 1993; Chiu, 2007).” (S. 238)</p> <p>- “in this study we intended to explore the possibility of the use of animations to diagnose students’ understanding of gas particles, because we considered that this type of assessment to be crucial in chemistry education.“ (S. 238 f.)</p> <p>- Präkonzept “empty space between the particles (Griffiths and Preston, 1992).” (S. 239)</p> <p>- Präkonzept “pre-service student teachers think that particles in gases are arranged in an orderly rather than a disorderly fashion (Gabel et al., 1987). (S. 239)</p> <p>- “Stavy (1990) found similarity between students’ and historical scientists’ conceptions of particles.” (S. 239)</p> <p>- “Building upon the work of Grossman et al. (1999), we conceptualize PCK for science teaching with five components: orientations toward science teaching; knowledge and beliefs about the science curriculum; knowledge and beliefs about the students’ understanding of specific science topics; knowledge and beliefs about assessment in science; and knowledge and beliefs about instructional</p>	<p>- “1. How do students perform on the test items at different grades?</p> <p>2. What are the characteristics of students’ mental models of the behavior of gas particles held by the participating eighth and ninth graders?</p> <p>a. What types of students’ mental models of the behavior of gas particles do eighth and ninth graders hold?</p> <p>b. How do students change their mental models when they face similar problems with different representations of gas particles?</p> <p>c. Do students hold a consistent model for explaining the behavior of gas particles?</p> <p>3. Are there any differences between how students perform in a test about the behavior of gas particles and science teachers’ predictions of how they will perform?” (S. 240)</p>	<p>- “(1) to develop a concept map on the characteristics and behavior of gas particles; (2) to establish propositional statements of key concepts that were related to the topics investigated in this study; (3) to interview students for their explanations, which were later transformed into multiple-choice test items; and (4) to develop a computerized test package that could dynamically reveal the characteristics of gas particles for students to select the most appropriate representations of their thinking about the behavior of gas particles.” (S. 240)</p> <p>- “Cronbachs Alpha = .76” (S. 241)</p>	<p>eighth graders (n = 102), ninth graders (n = 92), and physical science teachers (n = 31) in junior high schools in Taiwan.</p>	<p>Schüler, Schülerinnen und Lehrkräfte</p>

			<p>strategies for teaching science.” (S. 239)</p> <ul style="list-style-type: none"> - “Multiple representations provide diverse opportunities for students to construct the same conception from different perspectives.” (S. 240) - “Adadan’s (2010) study suggested that multi-representational instruction promotes conceptual change among students, and that teachers should integrate visual tools, such as animated or static particulate representations, in their teaching practice.” (S. 240) - Es existieren signifikante Durchschnittsunterschiede zwischen Acht- und Neuntklässlern. (S. 241) 				
42	2	Lin und Chiu (2010)	<ul style="list-style-type: none"> - „compare the characteristics and sources of students’ mental models of acids and bases with a teacher’s anticipations and, based on this comparison, to explore some possible explanations why motivated students might fail to learn from a subject-knowledgeable chemistry teacher.” (S. 1617) - “focuses on the mental models of three high achievers and three low achievers who were interviewed in depth.” (S. 1617) - Vier Präkonzepte zu Säuren und Basen entdeckt (S. 1617) - Lehrerin nahm Präkonzepte („mentale Modelle“) von guten SuS korrekt, von schlechten SuS inkorrekt wahr. (S. 1617) - Lehrerin verstärkte sogar Präkonzept der schlechten SuS. (S. 1617) - “10 different terms, such as misconceptions,” “alternative conceptions or framework,” or “children’s ideas,” that have been used to refer to students’ conceptions. For purposes of consistency, in discussing the findings, this paper uses “students’ conceptions” or “students’ preconceptions” to refer to conceptions that are not 	<ul style="list-style-type: none"> - “explore the role of the teacher’s PCK in influencing students’ conceptual learning.” (S. 1619) - “1. What are the differences between high and low achievers with regard to the changes and the sources of their mental models in acids/bases? 2. What are the differences between the students’ mental models and their chemistry teacher’s anticipations in terms of the changes in and sources of their mental models in acids/bases?” (S. 1623) 	<ul style="list-style-type: none"> - “The instrument comprised 34 two-tier questions on six topics: “the acidity or basicity of a solution,” “neutralization,” “diluting weak acids and bases,” “the acidity or basicity of dissolved gas solutions,” “ionized electrolytes,” and “strengths of acids and bases”.” (S. 1625) - “Reliability coefficient (KR-20): .8” (S. 1626) 	N = 1 Lehrerin, N = 38 Schülerinnen und Schüler	Chemielehrkräfte und v. a. Chemieschülerinnen und -schüler

			<p>scientifically correct.” (S. 1617)</p> <ul style="list-style-type: none"> - “teachers would prefer that the vast amount of research results regarding students’ preconceptions about science be compiled into an informative database that any teacher could readily use.” (S. 1618) - “the National Science Concept Learning Study (NSCLS), to investigate Taiwanese students’ science understanding and the sources that influence their learning (see the special issue of the <i>International Journal of Science Education</i>, 2007, Vol. 29(4)).” (S. 1618) - “Oversby (2000) points out, ‘many of the difficulties in chemistry appear to stem from a lack of explicit reason for using a restricted language’ (p. 239).” (S. 1619) - “two-tier diagnostic test on acids and bases before and after instruction and were observed during classroom instruction.” (S. 1624) 				
43	2	Nelson und Davis (2012)	<ul style="list-style-type: none"> - „Our findings suggest that with instruction, preservice elementary teachers can adopt criterion-based approaches to evaluating students’ scientific models.” (S. 1931) - Die Selbstwirksamkeitserwartung steigt, wenn hierin Wissen vermittelt wird. (S. 1931) - “In scientific modeling-based instructional approaches, students engage in many authentic scientific practices by creating and using representations of their understandings of science concepts in the form of scientific models.” (S. 1932) - “After they construct models, students use the models for communicative and sense-making purposes, evaluate the model’s fitness according to a set of criteria related to its purpose, and revise the model to better align with those evaluation criteria. Taken together, model construction, use, evaluation, 	<ul style="list-style-type: none"> - “(1) How do preservice elementary teachers evaluate elementary students’ scientific models? and (2) How do preservice elementary teachers’ knowledge, skills, and self-efficacy for model evaluation change during the elementary science teaching methods semester?” (S. 1933) 	Interviews und Auswertung schriftlicher Bewertung von Schülermodellen.	N = 4 (intensive Betrachtung) Lehramtsstudierende Bzw. N = 35 (oberflächliche Betrachtung)	Primary school teachers science models

			<p>and revision comprise the elements of modeling practice.” (S. 1932)</p> <p>- “Another key element of effective, model-centered science teaching is teacher self-efficacy (Bandura, 1997) for teaching using modeling-based pedagogies.” (S. 1933)</p> <p>- “PCK for scientific modeling (Schwarz, Meyer, & Sharma, 2007).” (S. 1934)</p>				
44	2	Oh und Kim (2013)	<p>- “study explored how science content knowledge was pedagogically transformed in Korean elementary classrooms.” (S. 1590)</p> <p>- “Korean teachers often engaged in transforming science content knowledge by means of different semiotic modes, including language, pictures, materials, actions, and their complex combinations.” (S. 1590)</p> <p>- “their representations of scientific knowledge were in diverse forms, such as personifications, analogies, quiz questions, pictorial models, diagrams, animations, real-life examples, hand demonstrations, videos, flash tools, and songs-and-dances.” (S. 1590)</p> <p>- “The multimodality of classroom representations of scientific knowledge reflects the nature of science that the development of scientific ideas is accompanied with the use of multiple semiotic resources. For instance, scientific models are constituted in verbal, visual, mathematical, and other semiotic resources so that they can organize a great deal of information and reasoning processes as visible and tangible (Gilbert & Boulter, 2000; Magnani, Nersessian & Thagard, 1999; Oh & Oh, 2011).” (S. 1594)</p> <p>- “the analysis framework involves five different semiotic modes used for representing scientific knowledge: verbal, pictorial, mate-</p>	<p>- “how do elementary teachers, who are often thought to have a weak knowledge base in science, use different modes of representation when they engage in pedagogical transformation of science content knowledge?” (S. 1595)</p>	<p>- “13 video-recordings of Korean elementary science classrooms were collected and transcribed as data.” (S. 1595)</p> <p>- “When more than one mode was utilised, the most distinctive mode was considered primary while other auxiliary modes, such as denoting words and incomplete line drawings, were not considered for frequency counting.” (S. 1597)</p> <p>- “After the representational modes were all coded, a concrete form of pedagogical transformation, such as an analogy, real-life example, demonstration, and animation, was identified for each case.” (S. 1597)</p>	N = 5 Grundschullehrkräfte (S. 1590)	Naturwissenschaften un- terrichtende Grund- schullehrkräfte

			rial, mathematical, and actional.” (S. 1596)				
45	2	Okanlawon (2010)	<p>- „this paper attempts to utilize those ideas on the components of pedagogical content knowledge (PCK) to formulate a framework to help clarify the knowledge that good chemistry teachers possess, and how that knowledge is applied in teaching stoichiometry.” (S. 27)</p> <p>- Stöchiometrie “There is a consensus among chemistry educators that learning to solve those problems require a good mastery of chemistry concept, ability to construct and balance chemical equations and using the balanced chemical equations to calculate the masses of chemical substances involved in the reactions.” (S. 28)</p> <p>- “Unfortunately, stoichiometry calculations have always been difficult for students [4-6]. The reasons for this difficulty lies in the complexity of conducting these calculations that require an understanding of the mole concept, constructing and balancing chemical equations, algebraic skills, and interpretation of a word problem into procedural steps that lead to the correct answer.” (S. 28)</p> <p>- “PCK for stoichiometry includes teachers' 'bag of tricks' and motivational 'tools' that can be used to develop in students better and more strategic problem solving techniques. It is in continuous growth and is kept constantly under review and reconstruction. It includes knowledge of how best to present stoichiometric concepts and ideas to students. It also includes knowledge of teaching strategies that could incorporate appropriate conceptual representations in order to address learners' difficulties and misconceptions regarding stoichiometry.” (S. 30)</p>	<p>- “1. Why teaching reaction stoichiometry? 2. What are the students' typical understanding and misunderstanding regarding reaction stoichiometry? 3. What curriculum materials would be most appropriate in the light of curriculum aims and objectives? 4. What strategies, representations, problem-solving models and illustrations would be most suitable for the students?” (S. 32)</p>	./.	./.	./.

		<p>- "Grossman: (a) knowledge and beliefs about the purpose of teaching a subject at different grade levels, (b) knowledge of students' understanding, conceptions and misconceptions of particular topics in a subject matter, (c) knowledge of curriculum materials available for teaching particular subject matter as well as knowledge about both the horizontal and vertical curricular for a subject and (d) knowledge of instructional strategies and representations for teaching particular topics [20]." (S. 32)</p> <p>- "objectives of teaching reaction stoichiometry are to enable the students to:</p> <ol style="list-style-type: none">1. define stoichiometry and distinguish between composition and reaction stoichiometry;2. identify the major types of reaction stoichiometry problems;3. predict the products of chemical reactions given the reactants;4. describe chemical reaction types and classify them;5. write balanced chemical equations for simple reactions;6. interpret balanced chemical equations in terms of interacting moles, representative particles, masses and gas volume at STP;7. distinguish between products and reactants in a chemical equation;8. determine the reactant-product (or product-reactant) pair that will provide the relevant stoichiometric relationship for calculation;9. construct mole ratios from balanced equations and apply these ratios in mole-mole, mole-mass, mass-mole and mass-mass stoichiometric calculations;10. calculate the amount of product expected to be formed in a chemical reaction, given the amounts of reactants used;11. calculate the amount(s) of				
--	--	--	--	--	--	--

		<p>reactants which need to be used in a chemical reaction in order to produce a specified amount of product;</p> <p>12. use stoichiometric methods to solve for composition of a two-component mixture in which only one of the components involved in the chemical reaction (e.g., an impure Na_2CO_3 treated with excess dilute hydrochloric acid solution);</p> <p>13. use stoichiometric methods to solve for composition of a two-component mixture in which the two components are involved in the chemical reaction (e.g., a mixture of Na_2SO_3 and MgSO_4 treated with excess BaCl_2 solution);</p> <p>14. use stoichiometric methods to identify the limiting reagent and excess reagent; and calculate the amount of expected products produced, and the amount of excess reagent left over upon completion of the reaction from a non-stoichiometric mixture of reactants;</p> <p>15. use stoichiometric methods to predict the theoretical yield and percent yield given the mass (number of moles) of each reactant and the actual yield of a reaction;</p> <p>16. calculate the mass (number of moles) of each reactant required given the percent yield and the mass (number of moles) of products desired.” (S. 33)</p> <p>- Präkonzepte:</p> <p>“a. equate the mass ratio of atoms in a molecule with the ratio of the number of these atoms, and the mass ratio with the molar mass ratio [31];</p> <p>b. calculate the molar mass of a given substance by summing up the atomic masses and then multiplying or dividing this sum by the coefficient of the substance in the chemical equation; others do not understand the significance of the coefficients in a chemical equation at all</p>				
--	--	--	--	--	--	--

			<p>[4];</p> <p>c. confuse the concepts of conservation of atoms and possible non-conservation of molecules or do not take into account the conservation of atoms or mass at all [30];</p> <p>d. cannot determine the 'limiting reagent' in a given problem, when one substance is added in excess [32];</p> <p>e. confuse or do not know the definitions of and relationships between stoichiometric entities in general [5];</p> <p>f. believe that one mole means the same as one particle [6].” (S. 34)</p>				
46	2	Padilla et al. (2008)	<p>- “Mortimer’s (1995) Conceptual Profile Model (CPM), defined as a ‘superindividual system of forms of thought’ (p. 270), which describes different thinking routes to one concept.” (S. 1390)</p> <p>- “perceptive/intuitive, empiricist, formalist, rationalist, and formal rationalist.” (S. 1396)</p>	<p>- “● What is the nature of university professors’ PCK about ‘amount of substance’?</p> <p>● What are these professors’ ways of thinking about this concept as explored through Mortimer’s CPM?” (S. 1393)</p>	<p>CPM: Mortimer’s (1995) Conceptual Profile Model, sowie die Fragebögen: CoRE</p> <p>“1. Why is it important for students to learn this idea and what you intend teaching it?</p> <p>2. From STS and historical context, why is it important for students to learn this?</p> <p>3. What else you know about history, philosophy, and epistemology of this idea?</p> <p>4. Difficulties/limitations connected with learning this idea.</p> <p>5. Difficulties/limitations connected with teaching this idea.</p> <p>6. Knowledge about students’ thinking that influences your teaching of this idea.</p> <p>7. Teaching procedures for engaging students with this idea (analogies, metaphors, examples, demonstrations, reformulations, etc.).</p> <p>8. Specific ways for ascertaining students’ understanding or confusion around this idea.” (S. 1395) und PaP-Ers</p>	N = 4 ProfessorInnen der Chemie	ChemieprofessorInnen
47	2	Papageorgiou et al. (2010)	<p>- Grundschullehrkräfte “similar misconceptions to pupils (Appleton, 1995; Baker, 1994; Harlen,</p>	<p>- “(1) What was the affect of the course on teachers’ understanding of the parti-</p>	<p>Chi-Quadrat-Tests pre post und jeweils Vergleich zu bisherigen Ergebnissen anderer Studien zu</p>	<p>- “Greek primary school teachers’ (n = 162)” (S. 629)</p>	<p>Grundschullehrkräfte für Naturwissenschaften (Chemie/Physik)</p>

			<p>Holroyd, & Byrne, 1995; Krugar, Summers, & Palacio, 1990; Mant & Summers, 1995; Smith & Neal, 1989; Summers, 1992; Summers & Krugar, 1992)” (S. 630)</p> <ul style="list-style-type: none"> - Fortbildungen für Grundschullehrkräfte sind notwendig. (S. 630) - Teilchenkonzept erst in Sek I „(Martin, Mullis, Gonzales, & Chrostowski, 2004)“ (S. 631) - Aber Aggregatzustände sind ohne dies nicht gut ermittelbar. (S. 631) - „In both preand post-tests, teachers’ particle ideas varied from the solid to gas states in a different way to younger pupils. The teachers’ resistance in accepting particle ideas is stronger for the solid state than the gas state; the difference is statistically significant [$\chi^2= 226, p < 0.000$].” 	<p>cle model and their explanations of physical phenomena during this course?” (S. 631)</p> <ul style="list-style-type: none"> - “(2) What is the relationship between the degree of teachers’ understanding of the particle model and their explanations of physical phenomena?” (S. 632) 	Präkonzepten von SuS		
48	2	Romine und Walter (2014)	<ul style="list-style-type: none"> - “only 4 of 10 Americans agree that humans and higher apes share a common ancestor (Newport, 2009). Among Americans with some college education, 49% accept the theory of evolution for plants and non-human animals, and only 22% agree that humans evolved from earlier species (Lovely & Kondrick, 2008).” (S. 2872 f.) - “It is critical to have valid and reliable instruments to accurately measure undergraduates’ knowledge as instructors and researchers examine their students’ evolution thinking and learning.” (S. 2873) - “Macroevolutionary concepts, in contrast, include the formation, radiation, and extinction of species and higher order natural groups comprising a most recent common ancestor and all its descendants over long periods of time (Campbell & Reece).” (S. 2873) - “instruments to measure knowledge of macroevolution 	<ul style="list-style-type: none"> - “(1) What is the dimensionality of the MUM? (2) Does the MUM show evidence for construct validity from the perspective of the Rasch model? (3) Does guessing damage the MUM’s construct validity? (4) Do the measurement characteristics of the MUM undergo significant changes over the course of a general education biology course?” (S. 2875) 	<p>- KTT: reliability of 0.82, IRT: Auswertungssoftware BIG-STEPSSoftware. (S. 2879)</p> <ul style="list-style-type: none"> - “Items with mean squares fit indices between 0.5 and 1.5 are generally considered productive for measurement of the underlying construct from the Rasch perspective (Wright & Linacre, 1996).” - “Reliability of person measures was 0.74 and 0.78 for the pre- and post-tests, respectively. Item measurement reliability was 0.98 for both tests.” (S. 2880) 	<p>“N = 315 Studierende: pre-test, N = 291 completed post-test. 270 students completed both pre- and post-tests.” (S. 2876)</p>	Biologiestudierende (Lehramt)

		<p>concepts were limited to the Tree Thinking Concept Inventory, an unpublished instrument assessing knowledge of phylogenetic trees (Naegle, 2009).” (S. 2873)</p> <ul style="list-style-type: none">- Entwickler “Measure of Understanding of Macroevolution ... Nadelson and Southerland (2010)” (S. 2873)- “This 27-item dichotomous multiple choice instrument was designed to measure five dimensions related to the understanding of macroevolution: deep time (5 items), phylogenetics (4 items), speciation (8 items), fossils (5 items), and the nature of science (5 items).” (S. 2873)- Validierung: Die Items wurden 20 EvolutionslehreexpertInnen und fünf FachwissenschaftlerInnen vorgelegt.- Scharfe Kritik hinsichtlich Konstruktvalidität durch Novick and Catley (2012): “students could get the answers correct without even looking at the question stem and proceeded to test this hypothesis.” (S. 2874)- “When guessing is not expected to be an issue, the guessing parameter can be constrained to zero. This leaves the two-parameter logistic (2PL) model, in which difficulty and discrimination parameters are allowed to vary for each item.” (S. 2878)- “The 2PL model can be further simplified to a one-parameter logistic (1PL) model by forcing the discrimination parameter to be the same for each item and adjusted to the value which achieves the best average fit over all items. A 1PL or 2PL model can be further simplified by constraining the discrimination parameter to be 1 for all items, thus defining the simple logistic model, also called the Rasch model			
--	--	---	--	--	--

			(de Ayala, 2009).” (S. 2878) - “The Rasch criterion can be expressed as follows: the probability of a student answering an item correctly should be proportional only to the difference between the student’s ability and the item’s difficulty (Wright & Stone, 1979).” (S. 2878)				
49	2	Schmelzing et al. (2013b)	- „The development of the test was based, among other considerations, on a review of research literature on PCK and an analysis of 50 videotaped biology lessons.” (S. 1369) - “In-service biology teachers scored higher on the test than preservice teachers (effect size Cohen’s d, 0.65) on one hand and, also, than biologists (Cohen’s d, 1.00) on the other hand.” (S. 1369) - “PCK implies two types of knowledge: declarative knowledge and procedural knowledge (Heller, Daehler, Shinohara & Kaskowitz, 2004).” (S. 1371) - “Declarative PCK or knowing that (Baumert, Blum & Neubrand, 2004; Ryle, 1971) has also been described in terms of PCK-on-action (Park & Oliver, 2008), theoretical-formal PCK (Fenstermacher, 1994), or propositional PCK (Knight, 2002). Declarative PCK is factual knowledge that can easily be expressed in sentences or indicative propositions (Anderson, 1981; Polanyi, 1958).” (S. 1371) - “the PCK test should cover two test scales: knowledge of student learning and conceptions (PCK I) and knowledge of representations and strategies (PCK II).” (S. 1376) - Videoanalyse: “aiming to identify frequently used models, representations, and explanations.” (S. 1376) - “Thus, the open-ended items measured the size of a PCK repertoire instead of distinguishing between “right” and “wrong” PCK	./.	“15 items distributed across 2 scales. Effektstärke “Cohens d” (S. 1369) - “The ICCunjust indicated excellent agreement of the two independent raters on the item scoring, ICCunjust=0.77, F(9,9)=7.50, p=0.003. Twenty-five items—which did not match the valid item difficulty, $0.20 < P_m < 0.80$, also showed a low discrimination coefficient (r_{it} , $rit < 0.30$, or were found to be ambiguous for solving or evaluating—were rejected. The final entire PCK test consisted of 15 open-ended items in 2 scales, one scale consists of 5 items and the other scale consists of 10 items.” (S. 1377) - Offene schriftliche Items (S. 1378) - “whole PCK test, $\alpha = 0.85$ ” (S. 1381) - “According to Shrout & Fleiss (1979), the interrater agreement for the whole PCK test was excellent, $7 ICCunjust = 0.92$, $F(299,299) = 24.45$, $p < 0.001$.”	- “93 preservice and in-service biology teachers and 12 biologists.” (S. 1369) - Pilot: “Forty-two participants (10 student teachers from the university and 32 in-service biology teachers)” (S. 1377) - Hauptstudie: “sample (n = 93) of student teachers (preservice teachers at the university, n = 22), trainee teachers (preservice teachers in teacher traineeship, n = 22), and experienced teachers (in-service teachers, n = 49).” (S. 1379) - Validierungsstudie: n = 12 abgeschlossene BiologInnen (S. 1379)	Biologielehrkräfte

			<p>responses (Baxter & Lederman, 1999; Kromrey & Renfrow, 1991).” (S. 1378)</p> <ul style="list-style-type: none"> - “The Kolmogorov-Smirnov test confirmed a normal distribution of test scores/test item difficulties for all three subsamples.” (S. 1381) - Dauer: 20 min (S. 1384) 				
50	2	Sorgo et al. (2014)	<ul style="list-style-type: none"> - „freshmen’s level of knowledge about genetics, evolution, human evolution, the nature of science.” (S. 6) - V. a. NoS mangelhaft (S. 6) - Weltweite Evolutionsstudien „showing that the results from one study cannot simply be transferred from one country to another”. (S. 7) - “The knowledge of nature and philosophy of science can be regarded as the key to the acceptance of evolution.” (S. 13) - “Questions on the knowledge on evolution were best answered by the Czech students and students from all other countries did not exceed the 50% of the scores on average.” 	<ul style="list-style-type: none"> - „to determine the differences in the a) level of knowledge about genetics, evolution, human evolution, and the nature of science; b) opinions on evolution, and c) the presence of non-scientific explanations among Czech, Slovakian, Slovenian and Turkish freshmen students.” (S. 6) 	<ul style="list-style-type: none"> - “Students answered the items by choosing one of three options: disagree (1), undecided (2) and agree (3).” (S. 7) - “• Religiosity was measured by one statement: ‘I would declare myself as a religious person’. • Actual school experiences on evolution teaching consisted of 4 items. • Active personal effort consisted of 4 items. • Opinions on how evolution should be taught consisted of 3 items. • Perceived importance of evolution consisted of 3 items.” (S. 7) - Pilot: “Based on answers of 204 secondary school students and a procedure proposed by Selwyn (1997) and Lavonen et al. (2004), the final pool of questions was compiled.” (S. 7) - Ausfüllzeit: 20 min/Test (S. 7) - “Kolmogorov-Smirnov test was first run to test the normal distribution.” (S. 7) 	<ul style="list-style-type: none"> - “994 first-year university students from the Czech Republic (276; 27.8%), Slovakia (212, 21.3%), Slovenia (217, 27.3%) and Turkey (235, 23.6%) participated in this study.” (S. 6) - “The larger number of them (708; 71.2%) were prospective teachers and 286 (28.8%) came from a non-educational background.” (S. 7) 	
51	2	Strübe et al. (2014)	<ul style="list-style-type: none"> - „However, there are only few professional knowledge tests analysing teachers’ pedagogical content knowledge of using models and chemistry language in chemistry classes.“ (S. 380) - “In this article pedagogical content knowledge is defined as the knowledge that enables teachers to structure, link, represent, and explain the content to students (Schmelzing, Wüsten, Sandmann 	./.	<ul style="list-style-type: none"> - “Two booklets, each containing 15 items (a set of PCK items of using models and chemistry language) were developed for the pilot study. All items are in a closed item-format, so teachers can rate on a 6-point Likert-scale.” (S. 383) - “Nine German professors of chemistry education at six universities (2 female, 6 males) and one teacher trainer (1 male) 	<p>Pilotmodelle Lehrkräfte: $n = 23$; ExpertInnenrating: $n = 9$ Expertinnen in Chemiedidaktik</p>	Chemielehrkräfte

			<p>and Neuhaus, 2010; Krauss, Neu-brand et al., 2008).”</p> <ul style="list-style-type: none"> - “Research on teachers’ pedagogical content knowledge on using models and modelling is rare in Germany.” (S. 381) - “For this reason, new PCK items focusing on using models and chemistry language in class have been developed. The items are included in a new PCK test called FEMo (German short form for chemistry language (Fachsprache), experiments and models) and were developed by the authors.” (S. 383) - “Each item has an item stem that describes a classroom situation, e.g. a presentation of a student drawing which represents a model or a model which should be used. Every situation is presented as if the teacher, who fills out the test, helps a novice teacher.” (S. 383) - “The items are subdivided into modelling processes; criticizing models, knowledge of models, and use of models.” (S. 383) - “Another problem is the high number of excluded relations of the model items. It reduces scoring for the items in a substantive way. For example only for three relations of item MMp 4 a scoring could be done. This reduces total scoring from 6 to 3 points for this item. In addition, scoring for every item is different. Some items just have a maximum number of 2 points whereas others have a maximum number of points of 6. Nevertheless, analysing model items by using relations is a good alternative to a right and false analysis, because acting in class is influenced by different aspects that is why there is no right or false behaviour in class.” (S. 386) 		<p>conducted the additional expert rating.“ (S. 384 f.)</p> <ul style="list-style-type: none"> -“The median of classified data is used as the so-called aggregate expert of the model items. After this, the four answers of each item were ranked (see Tab. 4 and 5). The rank of an answer describes the position in the order that was done after the ranking. The order was the basis for creating quasi-ranked pairs (Thillmann, 2008; Walpuski et al., 2012) (Tab. 4 and 5). For the fifteen model items fifteen orders and 84 quasi-ranked pairs (relations) were created (booklet A: 36 relations and one item is a true/false item; booklet B: 48 relations).” (S. 385) - “Cronbach’s alpha was calculated by using data, in which missing data were recoded into zero points as it is common for achievement tests.” (S. 385) - ”The expert-rating shows a good internal consistency of $\alpha = .81$.” - “The sum scores of both booklets are distributed normal which was tested by the Kolmogorov-Smirnov-Test.” (S. 386) - “Relations, which show a separation effect lower than .3 were revised by formulating the alternative answers much more clearly.” (S. 386) - “In this manner 12 items with 48 answer alternatives were devised. These 12 items, 12 model items and 14 experiments items were arranged in the new booklet FEMo.” (S. 386) 		
52	2	Unal et al. (2014)	- “mental models of “model” concept were determined and	./.	- „Semi-structured interviews that consisted of 7 questions	- “14 teacher educators” (S. 674) = N = 14	LehrkräfteausbilderInnen

			entitled with “ <i>optimum mental model</i> ”, “ <i>proximate mental model</i> ”, “ <i>goal oriented mental model</i> ” and “ <i>inconsistent mental model</i> ” by the researchers.” (S. 678)		<p>were used to collect data.” (S. 674)</p> <p>- “<i>Concept map: Models</i>” (S. 674)</p> <p>- “main interview questions were as follows:</p> <ol style="list-style-type: none"> 1. What does the concept of model mean to you? 2. What kind of characteristics do you believe must be available in a model? 3. Can you explain the modeling process? 4. What do you think is the objective of using models? 5. Which model types are you aware of? 6. Which models do you make use of in your courses? 7. Can you draw a concept map related to the “<i>model</i>” concept?” (S. 676) <p>- Gibt Conceptmap als Referenz (S. 677), eigene Conceptmap wird verglichen.</p> 		
53	2	Usak et al. (2011)	<p>- „The results revealed deficits in the subject matter knowledge of the student teachers. It also became obvious that the teachers in this sample held very traditional and teacher-centred beliefs when it came to chemistry teaching at the secondary level. Their teaching attitudes were geared mainly towards the acquisition of facts by pupils, and often ignored the development of process-oriented skills.” (S. 407)</p> <p>- “Much research has been performed in the past in order to better understand secondary students’ conceptions, misconceptions, and their learning difficulties concerning chemical reactions (Andersson 1986, 1990; BouJaoude 1991; Hesse and Anderson 1992; Johnson 1998, 2000a, 2000b, 2002; Eilks, Möllering, and Valanides 2007; Stavridou and Solomonidou 1998;</p>	<p>– “what is the SMK of beginning student teachers in this study’s sample concerning selected aspects of a basic understanding of chemical reactions? And what do the student teachers from this sample think concerning the teaching of chemical reactions in initial science lessons at the lower secondary level?” (S. 411)</p>	<p>- “A multiple choice test combined with open-ended explanations and semi-structured individual interviews were used as instruments for data collection (Bogdan and Biklen 1992).” (S. 412)</p>	<p>- “Thirty beginning science student teachers were tested, with an additional eight interviews from the student teachers in the same sample.” (S. 407)</p>	Chemielehrkräfte

			Solomonidou and Stavridou 2000).” (S. 408)				
54	2	Valanides und Angeli (2006)	<p>- „school curricula are changing in order to become more student centered, connect school subject matter to real-life authentic situations, and promote student understanding, conceptual change, and thinking rather than rote memorization or drill and practice.” (S. 87)</p> <p>- “information and communication technology (ICT) ICT can help students visualize difficult-to understand concepts, build models for facilitating understanding, and interact with specific parts of the learning environment to explore and test ideas.” (S. 87)</p> <p>- “Penner (2000/2001) argued that one method that could possibly assist the inquiry learning process is computer modeling.” (S. 88)</p> <p>- “Jonassen (2004) argued that the most ‘powerful method for engaging, fostering, and assessing conceptual change is the construction of qualitative and semi-quantitative models that represent their conceptual understanding of what learners are studying’ (p. 4).” (S. 88)</p> <p>- “Building explicit models externalizes internal mental models and gradually fosters conceptual change.” (S. 89)</p> <p>- “Comparing and evaluating models requires understanding that alternative models are possible and that the activity of modeling can be used for testing rival models’ (Jonassen, 2004, p. 5).” (S. 89)</p> <p>- “According to Bliss (1994), there are two types of modeling, namely, explorative modeling and expressive modeling. In explorative modeling, learners are asked to explore a readymade model that represents somebody else’s conceptions. Thus, in explorative modeling</p>	<p>-“1. Do preservice teachers’ models have a correct structure?</p> <p>2. How “real” are preservice teachers’ scientific models?</p> <p>3. What types of modeling experiences do preservice teachers infuse in their science lessons?” (S. 89)</p>	- Auswertung der Modellzeichnungen mit Model-It Software	N = 47 Lehramtsstudierende (S. 89)	Lehrkräfte

			<p>learners try out a model, look at cause-and-effect relationships, and draw conclusions based on the results of their exploration. They can also modify the model if there is a need to do so. In expressive modeling, learners express their own ideas and make a model or an external representation of their ideas. Subsequently, learners use their models to test hypotheses and, based on the results of their investigations, they improve their models. Morrison and Morgan (1999) argued that expressive modeling is much more productive for learning and conceptual change than is explorative modeling. 'We do not learn much from looking at a model – we learn a lot more from building the model' (p. 11).” (S. 89)</p>				
55	2	Van Dijk (2009)	<ul style="list-style-type: none"> - „aims to obtain an impression of teachers’ knowledge and beliefs concerning teaching evolutionary theory.” (S. 259) - “Educational Reconstruction for Teacher Education model (ERTE).“ (S. 259) - “The PCK-study shows that teachers’ attitudes toward students’ conceptions of evolutionary theory are not always constructive and that teachers often lack awareness of the historical nature of biology.” (S. 259) - ““Although teachers have some knowledge about students’ difficulties, they commonly lack important knowledge necessary to help students overcome those difficulties’ (Magnusson, Krajcik & Borko, 1999, p. 106).” (S. 259) 	<ul style="list-style-type: none"> - „(RQ1) What ‘subject matter knowledge for teaching’ do biology teachers have concerning the topic of evolutionary theory? (RQ2) What knowledge and beliefs do biology teachers have of students’ pre-scientific conceptions with respect to evolutionary theory? (RQ3) What knowledge and beliefs do biology teachers have of subject matter representations regarding the theory of evolution?” (S. 260) 	Qualitative Inhaltsanalyse (Mayering 2003) (S. 263)	N = 9 Biologielehrkräfte des Gymnasiums in Niedersachsen	Biologielehrkräfte
56	2	J. H. van Driel et al. (2002)	- „The study focused on PCK with respect to a central issue in science teaching, that is, the relation between observable phenomena, like chemical reactions, and macroscopic properties (e.g., boiling point, solubility) on the one hand, and	- “What development of the preservice teachers’ PCK can be identified and what is the influence of specific factors (i.e., teaching experience, institutional workshops, and the mentor) on	- “two written questionnaires, interviews with each preservice teacher and their respective mentors, and an audio recording of a specific workshop session in the teacher education program.” (S. 572)	N = 12 künftige Chemielehrkräfte, die bereits M.Sc. haben.	Chemielehrkräfte

			<p>their interpretation in terms of corpuscular characteristics on the other hand (<i>macro-micro</i>)." (S. 572)</p> <p>- "For secondary school students, shifting mentally between the macro and micro levels is usually problematic, whereas their teachers are often unaware of students' learning difficulties in this domain."</p> <p>- "we have tried to identify the influence of certain components of the preservice teacher education program (viz., specific workshops, student teaching experiences, and feedback from mentors) on this development." (S. 573)</p> <p>- "The <i>macro-micro</i> perspective concerns the relation between observable phenomena, like chemical reactions, and macroscopic properties (e.g., boiling point, solubility) on the one hand, and their interpretation in terms of corpuscular characteristics on the other hand (De Vos & Verdonk, 1996)." (S. 576)</p>	this development?" (S. 576)			
57	2	Wang et al. (2014)	<p>- „(1) Chemistry teachers' knowledge of some known chemistry models was limited; (2) Chemistry teachers preferred those models that were vivid when they chose models; (3) Teachers' modeling process was incomplete; (4) Teachers adopted a general pattern when applying models in chemistry teaching." (S. 211)</p> <p>- "The process of developing models modeling is central to scientists' daily practice; thinking and reasoning with models enables scientists to visualize the abstract processes and entities they are investigating, to provide explanations for them and to make predictions about them (Gilbert et al. 2000)." (S. 211)</p>	<p>- "(1) What is chemistry teachers' knowledge of known model? (2) How do chemistry teachers select model? (3) How do chemistry teachers construct model? (4) How do chemistry teachers use model in teaching?" (S. 212)</p>	<p>- "Item 1 was based on the model of graphite and examined chemistry teachers' knowledge of known models; Item 1 also examined how teachers would apply the graphite model in teaching; Item 2 presented four different pictures to reflect the principle of the petroleum fractionation in order to explore teachers' criteria in selecting models; Item 3 was based on the atomic planetary model to investigate how chemistry teachers construct models." (S. 217)</p> <p>- Auszählen und prozentuale Auswertung der Antworten</p>	N = 50 Chemielehrkräfte	Chemielehrkräfte

			<p>- Ausgangsfrage für Review: "How good is teachers' knowledge of model? How do teachers apply model in classroom?" (S. 212)</p> <p>- "quite a few researches focus on teachers' application of models in classroom." (S. 214)</p> <p>- "Glynn (1991) put forward the Teaching-with-Analogies Model, and it includes 6 steps:</p> <ol style="list-style-type: none">(1) introduce the target concept;(2) remind students of what they know of the analogy concept;(3) identify relevant features of the concept and analogy;(4) connect the relevant features;(5) indicate where the analogy breaks down;(6) draw a conclusion about the concept." (S. 214)			
--	--	--	---	--	--	--

9.2.3 Tabelle mit Zentralbegriffen und den zugeordneten Unterrichtsfächern

		Quelle	Zentralbegriffe	Unterrichtsfach
1	3	Beerenwinkel und Parchmann (2008)	- Umgang mit submikroskopischer Ebene - Präkonzepte - Conceptual Change - Verständnis über die Natur von Modellen und ihrer Rolle in den Naturwissenschaften	Chemie
2	3	Benedict und Bolte (2009)	- Umgang mit submikroskopischer Ebene/Diskontinuum - Teilchenstruktur - Präkonzepten rechtzeitig entgegenwirken - Conceptual Change	Chemie
3	3	Bindernagel und Eilks (2009)	- Teilchen	Chemie
4	3	Bühler und Erb (2010)	- Einfluss von Weltbild auf Kontinuums- und Diskontinuumsvorstellungen	Chemie/Physik
5	3	Christen (1994)	- Modelle adäquat an Niveaus anpassen können	Chemie
6	3	Gad und Mittelsten Scheid (2008)	- Planspiel als Modell eines komplexen Sachverhalts, der vereinfacht werden muss, jedoch nicht zu sehr, um der Wirklichkeit weitestgehend Rechnung zu tragen: -> Modelle weiterentwickeln können/Modelle adäquat einsetzen können.	Biologie
7	3	Graulich et al. (2010)	- Chemische Probleme intuitiv lösen	Chemie
8	3	Gröger (2010)	- Suche nach passender Analogie	Biologie/Chemie
9	3	Hammer (1988)	- Modelle adäquat einsetzen (vgl. Zitat S. 35 oben)	Chemie
10	3	Hilgers et al. (1994)	- Versuche als Modell für einen Vorgang im Körper nutzen -> Modelle weiterentwickeln können/Modelle adäquat einsetzen können.	Biologie
11	3	Keune und Daemmgen (1988)	- Glucose-Modell mit Zeichenkarton nachbauen: -> Modelle (weiter)entwickeln können	Biologie/Chemie
12	3	Klinger (2001)	- Modelle adäquat einsetzen können: Veranschaulichung vor Mathematisierung	Physik/relevant auch für Chemie
13	3	König und Reiners (2004)	- Dynamische 3D-computerbasierte Atom(-Bindungs-)darstellung bringt Erfolge: -> Modelle weiterentwickeln können/Modelle adäquat einsetzen können	Chemie
14	3	Leisner und Mikelskis (2004)	- Metakonzeptuelle Kompetenz hinsichtlich -> „Conceptual Change“ (S.120) durch „Modellmethode“	Physik/Chemie
15	3	Lethmate und Arning (2003)	- Modellmethode als Teil von -> Modelle adäquat einsetzen können	Geografie und Chemie bzw. Biologie
16	3	Litsche (1985)	- Gedankenmodelle -> adäquat einsetzen und weiterentwickeln können (vgl. S. 263)	Biologie
17	3	Markert (1988)	- Geeignete Synapsenmodelle adäquat einsetzen	Biologie
18	3	Meyer (1990)	- Modelle adäquat einsetzen und weiterentwickeln können	Biologie
19	3	Mikelskis-Seifert (2004)	- „Entwicklung metakonzeptueller Denkweisen ... im submikroskopischen Bereich“ (S. 18) - (Dis-)Kontinuum	Physik/Chemie

			<ul style="list-style-type: none"> - Modellmethode - Modelle weiterentwickeln können - Modelle adäquat einsetzen können - Präkonzepte - Conceptual Change 	
20	3	Mikelskis-Seifert und Leisner (2004)	<ul style="list-style-type: none"> - (Dis-)Kontinuum - Modellmethode - Modelle weiterentwickeln können - Modelle adäquat einsetzen können - Präkonzepte - Conceptual Change 	Physik/Chemie
21	3	Mikelskis-Seifert et al. (2005)	<ul style="list-style-type: none"> - Dynamisches (vgl. S. 36) Modellieren hinsichtlich - (Dis-)Kontinuum - Modellmethode - Modelle weiterentwickeln können - Modelle adäquat einsetzen können - Präkonzepte Conceptual Change 	Physik/Chemie
22	3	Nocke (2001)	<ul style="list-style-type: none"> - Modelle adäquat einsetzen 	Biologie
23	3	Rueckl und Ebinghaus (1989)	<ul style="list-style-type: none"> - Modelle adäquat einsetzen 	Physik/Chemie
24	3	Schneider (1981)	<ul style="list-style-type: none"> - Modelle weiterentwickeln können - Modelle adäquat einsetzen können 	Biologie
25	3	Steinbeck und Markert (1989)	<ul style="list-style-type: none"> - Modelle adäquat einsetzen können 	Biologie
26	3	Streller und Bolte (2008)	<ul style="list-style-type: none"> - Modelle weiterentwickeln können - Modelle adäquat einsetzen können 	Chemie/Geografie/Physik
27	3	Touché (1989)	<ul style="list-style-type: none"> - Teilchen- vs. Atommodelle - Modelle adäquat einsetzen können 	Chemie/Physik
28	3	Winnenburg (2000)	<ul style="list-style-type: none"> - Conceptual Change - Modelle adäquat einsetzen können - Mondphasen-Präkonzepte 	Physik/Astronomie/Chemie
29	3	Cohen und Yarden (2009)	<ul style="list-style-type: none"> - Makro-mikro-Beziehungen (131), auch bezogen auf (Dis-)Kontinuum hinsichtlich Zellen, Modelle (hier Zellen) sollten als Spiralcurriculum vermittelt werden (Zellen zumindest) - Präkonzepte (S. 145) Conceptual Change und Modelle adäquat einsetzen können hinsichtlich Zellen, aber auch hinsichtlich Lehrkräftebildung 	Biologie, auch hinsichtlich Chemie
30	3	Crawford und Cullin (2004)	<ul style="list-style-type: none"> - Dynamische Modelle-> Modelle unterscheiden, Präkonzepte und Conceptual Change, NoS hinsichtlich Modellen, Modelle adäquat einsetzen und weiterentwickeln können 	Biologie, offiziell Science
31	3	Drechsler und van Driel (2008)	<ul style="list-style-type: none"> - Präkonzepte, Modelle adäquat einsetzen, Mikro-/Makroorientierung (Dis-)vs Kontinuum 	Chemie
32	3	Eilks et al. (2009a)	<ul style="list-style-type: none"> - "Chemistry necessarily tries to explain macroscopic properties and phenomena on the sub-microscopic level." (S. 51) -> (Dis-)Kontinuum - Präkonzepte und Conceptual Change (Posner et al. 1982) (S. 51 f.) - Modellmethode: – "the model nature of the chosen explanation" (S. 54) 	Chemie, aber auch Physik

33	3	Garritz (2013)	- "Nature of Science will be investigated through history and philosophy of physical sciences (HPS)"(...) "describe an abstract atomic world far away from daily experience" -> (Dis-)Kontinuum (S. 1788)	Physik, auch hinsichtlich Chemie
34	3	Harrison und Jong (2005)	- "better understand the submicroscopic world by helping the students visualize particle interactions." (S. "2") -> (Dis-)Kontinuum, - "check students' understanding of the analogies by asking them to retell the links between a particular analog and target in their own words." (S. "2") -> Präkonzepte, -> Conceptual Change, Modelle adäquat einsetzen - "attention to limitations of the analogies by inviting the students to find where the analogies break down." (S. "2") -> Modellmethode	Chemie
35	3	Jong und van Driel (2004)	(Dis-)Kontinuum + symbolische Ebene , Präkonzepte	Chemie
36	3	Jong et al. (2005b)	Teilchen-Konzept, Präkonzepte	Chemie
37	3	Justi und van Driel (2005a)	Modellunterschiedsverständnis (CK), (Dis-)kontinuum	Chemie
38	3	Justi und van Driel (2005c)	- Modellmethode: "continuous process of producing and revising models." (S. 550) - Modellverständnis (CK, Curriculum Knowledge, PCK) (S. 549) - CK: Wissen über Modelle bzw. Wissen über den Modellierungsprozess vgl. Justi und Gilbert im Druck (S. 555) - Curriculum Knowledge: Umgang mit und Anpassung von curricularen Modellen (historische oder wissenschaftliche Modelle angepasst an Schule) und Einführung des Modellierens (S. 555) - PCK: "3.1 <i>Teaching models – purposes of their use</i> : the main purposes of the use of teaching models by either teachers or students using teaching models. 3.2 <i>Teaching models – production</i> : the nature of the models employed in their production and points that the teacher should take into account in producing different kinds of teaching models (all these points were discussed in our meetings). 3.3 <i>Teaching models – use in science teaching</i> : ways in which the notion of teaching models is deployed by the teacher. 3.4 <i>Conducting of modelling activities in science teaching</i> : teachers' role, characteristics of the discussion of students' models and teachers' previous experience. 3.5 <i>Knowledge of students' ideas about models and modelling</i> : status of this teacher knowledge." (S. 555) - 3.1 Modelle adäquat einsetzen - 3.2 Modelle adäquat einsetzen - 3.3 Modelle adäquat einsetzen - 3.4 Ausführung von Modellierungsaktivitäten - 3.5 Präkonzepte	Chemie (evtl. mit Physikanteil)
39	3	Justi und van Driel (2005b)	- Modellmethode - Modelle adäquat einsetzen - Modellverständnis	Chemie

			- Ausführung von Modellierungsaktivitäten - Präkonzepte	
40	3	Khourey-Bowers und Fenk (2009)	- (Dis-)Kontinuum („makroskopische zu abstrakten Modellen...“, S. 437) - Präkonzepte (441) - Conceptual Change (S. 441) - Teilchen (S. 442)	Chemie
41	3	Liang et al. (2011)	- Teilchen - Submikroskopisch (Dis-)Kontinuum (S. 238) - Präkonzepte (S. 238) - Conceptual Change (S. 238)	Physik/Chemieunterricht
42	3	Lin und Chiu (2010)	- Präkonzepte (hier: mentale Modelle) (S. 1617) - Conceptual Change (Taber 2001a) (S. 1618) - (Dis)Kontinuum/symbolisch (Johnstone)	Chemie
43	3	Nelson und Davis (2012)	- Modelle (weiter)entwickeln können (S. 1932) - Modellmethode („revise“) (S. 1932) - Bewerten von Modellqualität von SuS (=>Modelle weiterentwickeln), das heißt: Klarheit, Bedeutung, Übereinstimmung mit empirischen Belegen (Schwarz et al., 2009). - PCK for scientific modelling (Schwarz, Meyer, & Sharma, 2007)	Chemie/Biologie/Astrologie
44	3	Oh und Kim (2013)	- Modelle (Repräsentationen und Analogien) adäquat einsetzen	Chemie, Biologie, Physik (gleichwertig Science)
45	3	Okanlawon (2010)	- Modelle adäquat einsetzen können	Chemie
46	3	Padilla et al. (2008)	- Modelle adäquat einsetzen können nach Conceptual Profile Model	Chemie
47	3	Papageorgiou et al. (2010)	- Präkonzepte, Modelle adäquat einsetzen, Conceptual Change - Atome vs. Teilchen	Chemie/Physik
48	3	Romine und Walter (2014)	- Modelle adäquat einsetzen	Biologie
49	3	Schmelzing et al. (2013b)	- Wissen über (Repräsentationen und) Modelle -> Modelle unterscheiden können	Biologie
50	3	Šorgo et al. (2014)	- Modelle adäquat einsetzen	Biologie
51	3	Strübe et al. (2014)	- Modellierprozess, Modellkritik -> Modellmethode, Wissen über Modelle, Gebrauch von Modellen -> Modelle adäquat einsetzen können	Chemie
52	3	Unal et al. (2014)	- Präkonzepte (hier: mentale Modelle), Modellwissen	Naturwissenschaften, v.a. Chemie
53	3	Usak et al. (2011)	- Modelle adäquat einsetzen	Chemie
54	3	Valanides und Angeli (2006)	- Modelle adäquat einsetzen - Präkonzepte -> hier: mentale Modelle - Conceptual Change, - Modellieren (exploratives und expressives) - Modellmethode	Biologie
55	3	Van Dijk (2009)	- Präkonzepte und Conceptual Change (S. 259)	Biologie
56	3	J. H. van Driel et al. (2002)	- Makro/mikro = (Dis-)Kontinuum (S. 572)	Chemie
57	3	Wang et al. (2014)	-“symbolic, macroscopic, and submicroscopic.” (S. 212), d.h. (Dis-)Kontinuum	Chemie

9.2.4 Tabelle zu den quantitativen Studien und deren statistischen Schwächen, falls keine Vergleichsgruppe, keine Stichprobe und keine Randomisierungs- oder Auswertungsverfahren genannt wurden

		Quelle	Keine Vergleichsgruppe (Wartekontrollgruppe...)	N nicht oder nur ca. genannt	Auswertungsverfahren nicht genannt	Keine Randomisierung der Gruppen etc.
1	4	Bühler und Erb (2010)	./.	Nein, $N = 69$ (Elf- und Zwölftklässler)	Ja (wahrscheinlich Häufigkeitsverteilung)	Ja, nur eine Schule: Burg-Gymnasium Schorndorf
13	4	König und Reiners (2004)	Wartekontrollgruppe fehlt	Ja, Stichprobengröße fehlt, sowohl für die 9. wie für die 10. Klassen	Nur vage: „Die Anzahl der Lernenden wird gegen die erbrachte Leistung der Lernenden aufgetragen.“ (S. 335)	Ja, zumindest wird keine erwähnt
14	4	Leisner und Mikelskis (2004)	Wartekontrollgruppe fehlt	Ja, Stichprobengröße fehlt, sowohl für die 7., 8., 9. und 10. Klasse	Ja, es wird nur auf das „Untersuchungsdesign siehe Leisner & Mikelskis 2003, S. 140“ verwiesen. (S. 121)	Ja, zumindest wird keine erwähnt
19	4	Mikelskis-Seifert (2004)	Vergleichs- und Wartekontrollgruppe fehlt	Ja, ca. $N = 120$ (S. 19)	- Nein, wird nicht genannt: „Zur Auswertung der mit dem Fragebogen und der Concept Map erhobenen Testdaten wurden unterschiedliche Analysemethoden wie zum Beispiel Varianz- und Korrelationsanalysen, Kausalanalysen mithilfe von LISREL, LatentClass-Analysen mit WINMIRA sowie Auswertungen auf der Basis von Modalnetzen eingesetzt.“ (S. 20)	Nein, da keine Vergleichsgruppe vorhanden
20	4	Mikelskis-Seifert und Leisner (2004)	Vergleichs- und Wartekontrollgruppe fehlt	Ja, ca. $N = 120$ (S. 124)	- „Mit den gleichen Instrumenten erfolgte der Nachtest, der direkt im Anschluss an die Projektwoche stattfand.“ (S. 124): - Umgang mit Testwiederholungseffekten nicht beachtet: evtl. Lernen über den Test	Nein, da keine Vergleichsgruppe vorhanden
29	4	Cohen und Yarden (2009)	./.	Nein, alle Ns genannt	Obwohl das Scoring von Schülerantworten ein quantitatives Verfahren anregt, sind hierzu nur qualitative Auswertungen zu finden.	Convenient sample (Mayering) ist fraglich
40	4	Khourey-Bowers und Fenk (2009)	./.	Nein, alle Ns genannt, aber kleines N	Doch: t-Test	Nein, scheint randomisiert (LehrerInnen aus Stadt und Land...)
41	4	Liang et al. (2011)	Nein, gerade noch o. k., nur bei Lehrkräften ($N = 31$) zu klein, deswegen nicht ausgewertet (sauber argumentiert)	Nein, alle Ns genannt	Doch: t-Test	JA, Daten nur SuS aus einer großen Stadt! Nicht randomisiert
47	4	Papageorgiou et al. (2010)	Ja, keine Vergleichsgruppe (keine Wartekontrollgruppe ohne Intervention)	Nein, alle Ns genannt	Doch: Chi-Quadrat-Test	JA, Daten nur SuS aus einer großen Stadt! Nicht randomisiert

48	4	Romine und Walter (2014)	Wartekontrollgruppe fehlt	Nein, alle Ns genannt	Doch: Raschskalierung	Erhebung unverändert an nur eine Uni = nicht randomisiert
49	4	Schmelzing et al. (2013b)	Nein, gerade noch akzeptabel ($N = 20$)	Nein, alle Ns detailliert genannt	Doch: komplexe KTT: ANOVA usw.	Randomisierte Lehrkräfteauswahl, aber nur Niedersachsen.
50	4	Šorgo et al. (2014)	Nein, große Vergleichsgruppen international	Nein, alle Ns detailliert genannt	Doch: t-Test	Ja, nicht randomisiert.
51	4	Strübe et al. (2014)	Vergleichsgruppe fehlt (Pilot)	Nein, Ns genannt ($N = 23$ Chemielehrkräfte)	Doch	Randomisierung nicht genannt (Pilot)