

Studien zum Physik- und Chemielernen

M. Hopf, H. Niedderer, M. Ropohl, E. Sumfleth [Hrsg.]

311

Marvin Rost

Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht der Sekundarstufe I

Entwicklung und quantitative
Dimensionalitätsanalyse eines Testinstruments
aus epistemologischer Perspektive

λογος

Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf, Hans Niedderer, Mathias Ropohl und Elke Sumfleth

Diese Reihe im Logos Verlag Berlin lädt Forscherinnen und Forscher ein, ihre neuen wissenschaftlichen Studien zum Physik- und Chemielernen im Kontext einer Vielzahl von bereits erschienenen Arbeiten zu quantitativen und qualitativen empirischen Untersuchungen sowie evaluativ begleiteten Konzeptionsentwicklungen zu veröffentlichen. Die in den bisherigen Studien erfassten Themen und Inhalte spiegeln das breite Spektrum der Einflussfaktoren wider, die in den Lehr- und Lernprozessen in Schule und Hochschule wirksam sind.

Die Herausgeber hoffen, mit der Förderung von Publikationen, die sich mit dem Physik- und Chemielernen befassen, einen Beitrag zur weiteren Stabilisierung der physik- und chemiedidaktischen Forschung und zur Verbesserung eines an den Ergebnissen fachdidaktischer Forschung orientierten Unterrichts in den beiden Fächern zu leisten.

Martin Hopf, Hans Niedderer, Mathias Ropohl und Elke Sumfleth

Studien zum Physik- und Chemielernen

Band 311

Marvin Rost

**Modelle als Mittel der
Erkenntnisgewinnung im Chemieunterricht
der Sekundarstufe I**

Entwicklung und quantitative Dimensionalitätsanalyse
eines Testinstruments aus epistemologischer
Perspektive

Logos Verlag Berlin



Studien zum Physik- und Chemielernen

Martin Hopf, Hans Niedderer, Mathias Ropohl und Elke Sumfleth [Hrsg.]

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Dissertation an der mathematisch-naturwissenschaftlichen Fakultät der Humboldt-Universität zu Berlin (2020)



© Copyright Logos Verlag Berlin GmbH 2021

Alle Rechte vorbehalten.

ISBN 978-3-8325-5256-5
ISSN 1614-8967

Logos Verlag Berlin GmbH
Georg-Knorr-Str. 4, Geb. 10
D-12681 Berlin

Tel.: +49 (0)30 / 42 85 10 90
Fax: +49 (0)30 / 42 85 10 92
<https://www.logos-verlag.de>

Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht der Sekundarstufe I

Entwicklung und quantitative Dimensionalitätsanalyse eines
Testinstruments aus epistemologischer Perspektive

Dissertation zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach: Chemie
Spezialisierung: Didaktik der Chemie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
von
M. Ed. Marvin Rost, geboren am 02.06.1986 in Berlin

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. habil. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät
Prof. Dr. Elmar Kulke

Gutachter/innen:
1. Prof. Dr. Rüdiger Tiemann
2. Prof. Dr. Jenna Koenen
3. Prof. Dr. Andreas Nehring

Tag der mündlichen Prüfung: 24.11.2020

Zusammenfassung

Das Nutzen von Modellen ist ein unverzichtbarer Bestandteil in der Chemie und in chemiebezogenen Lehr-/Lernumgebungen. In vielen Bildungsstandards auf der ganzen Welt ist das Nutzen von Modellen als hypothetisch-deduktives Verfahren, d. h. einer bestimmten Art der Erkenntnisgewinnung, vorgesehen. Gleichzeitig ist die empirische Befundlage zur Modellnutzung dabei sowohl für das Fach Chemie im Speziellen als auch für Naturwissenschaften im Allgemeinen uneinheitlich und enthält zahlreiche konzeptuelle Widersprüche. Die vorliegende Studie untersucht dieses Spannungsfeld, um einen Beitrag zur Strukturaufklärung zu leisten und Implikationen für einen weniger heterogenen Umgang mit Modellen abzuleiten.

Zu diesem Zweck wurden erkenntnistheoretische Rahmungen für Modelle in den Naturwissenschaften aufgearbeitet und auf Aspekte untersucht, die einerseits als Orientierungshilfen für die Konstruktion eines empirischen Instruments dienen und andererseits reflexiv zur Interpretation der erhobenen Daten eingesetzt wurden. In der Folge der Theoriearbeit wurden Modelle als erkenntnistheoretische Werkzeuge definiert, bei denen eher die situativen Zweckbestimmungen und weniger die ontologischen Definitionsbedingungen im Vordergrund stehen. Bei dem Instrument handelt es sich um einen Multiple-Choice-Fragebogen, der Schülerinnen und Schülern der 10. Jahrgangsstufe ($N = 524$) in Berliner Gymnasien zur Bearbeitung vorgelegt wurde. Die Fragen waren als Leistungsaufgaben konzipiert, bei der Modelle in chemiebezogenen Problemstellungen zum Durchlaufen eines hypothetisch-deduktiven Dreischritts (1.: Frage/Hypothese, 2.: Planung/Durchführung, 3.: Auswertung/Reflexion) genutzt werden sollten. Dazu dienten vier theoretisch hergeleitete Komponenten (kleinste Sinneinheiten, Relationen, Operationen, Regeln) als Strukturelemente jedes Modells. Die erhobenen Daten wurden mit Hilfe der Item Response Theorie ausgewertet und die so erhaltenen uni- und multidimensionalen Skalierungen statistisch miteinander verglichen.

Die Auswertungen zeigen, dass die Annahme einer 4-dimensionalen Skalierung im Sinne hierarchisch geschachtelter Modellkomponenten (sog. Within-Item-Dimensionality), inkl. eines psychometrischen Hintergrundmodells aus kognitiven Fähigkeiten, Fachwissen und Leseverstehen, die Antwortmuster besser erklären als andere, theoretisch ebenfalls plausible Strukturen.

Die erhaltene Inferenz auf das am besten passende Modell wurde an den erkenntnistheoretischen Theorierahmungen diskutiert und mit bereits vorhandenen empirischen Befunden ins Verhältnis gesetzt. Neben Vorschlägen zu Anschlussuntersuchungen, z. B. zum Argumentieren in modellbasierten Lehr-/Lernumgebungen, wird die Explikation der epistemologischen Vorannahmen beim Nutzen von Modellen als wichtigste Schlussfolgerung aus der Studie bewertet.

Abstract

Using models is an essential part in chemistry and chemistry education. It is conceptualised as a hypothetic-deductional framework in educational standards all over the world. From an empirical perspective, however, there are numerous conceptual discrepancies in both using models in chemistry, as well as in sciences in general. This study investigates the problematic trade-offs in this area of research and proposes a more generalisable structure of using models to lower the carved-out heterogeneity.

To that end, a review of epistemological frameworks was conducted. The theoretical results were used to construct an empirical instrument and to interpret the collected data. Models were defined as epistemic tools, which are rather defined by their subjectively intended purpose and less by their ontological conditions. The instrument itself is a multiple-choice questionnaire for 10th-degree high school students ($N = 524$) in Berlin, Germany. The questionnaire mainly consisted of problem-oriented tasks in chemistry and the students were supposed to work through a three-step conceptualisation of hypothetic-deductional reasoning (1.: hypothesis/question, 2.: planning/conducting, 3.: analysis/reflexion). Four theoretically derived model components (smallest entities, relations, operations, rules) were used as structural elements for all of the presented models. The collected data sets were analysed under the Item Response Theory paradigm. The resulting uni- and multidimensional scalings were statistically compared to each other.

The results favoured the assumption of a 4-dimensional scaling that consists of hierarchically nested model components (within-item-dimensionality), including a psychometric background model with general cognitive abilities, content knowledge and reading comprehension.

The 4-dimensional solution was discussed with respect to the epistemological framework and put into perspective with established empirical findings in the field. The main conclusions are, a) proposing follow-up studies regarding the relation between argumentation and scientific reasoning with models, and b) the necessity to explicitly state the particular epistemological assumptions when using models in chemistry and chemistry education.

Abbildungsverzeichnis

2.1.	Modell des Modellseins nach Mahr (2015).	29
2.2.	So genannte <i>Skelettformel</i> von Hexan-1-ol.	46
2.3.	Strukturformel von Adrenalin mit räumlicher Auszeichnung der OH-Gruppe.	46
2.4.	Exemplarische Darstellung des 7-dimensionalen Modells.	53
2.5.	Exemplarische Darstellung des 3-dimensionalen Modells.	53
2.6.	Exemplarische Darstellung des 4-dimensionalen Modells.	54
2.7.	Exemplarische Darstellung des 1-dimensionalen Modells.	55
4.1.	Schematische Darstellung der Datenerhebung.	60
4.2.	Konstruktionsschema für eine Aufgabe zu <i>Frage/Hypothese</i> mit <i>Relationen</i> beim Thema <i>Destillation</i> .	65
4.3.	Beispielaufgabe zu <i>Frage/Hypothese</i> mit <i>Relationen</i> beim Thema <i>Destillation</i> .	66
4.4.	ICC-Beispiel mit Misfit.	81
4.5.	ICC-Beispiel mit guter Passung.	82
5.1.	Grafische Verteilungsprüfung der Fähigkeitsschätzer im KFT der Vorstudie.	92
5.2.	Schulvergleich der KFT-Fähigkeitsschätzer in der Vorstudie.	92
5.3.	Regressionslinien für EAP-PVs der Modellaufgaben über den Fähigkeitsschätzern kognitiver Fähigkeiten getrennt nach Schulen.	94
5.4.	ICC der Aufgabe zu <i>Destillation</i> mit Denkschritt <i>Auswertung / Reflexion</i> und Modellkomponente <i>Regeln</i> .	95
5.5.	ICC der Aufgabe zu <i>Reaktionsgleichungen</i> mit Denkschritt <i>Planung / Durchführung</i> und Modellkomponente <i>Relationen</i> .	96
5.6.	Schwierigkeitsverteilung aus der 4-dimensionalen Aufgabenskalierung nach Themen.	105
5.7.	Schwierigkeitsvergleiche aus der 4-dimensionalen Aufgabenskalierung nach Themen.	106
6.1.	Überblick der Facetten von möglichen Anschlussuntersuchungen.	117
C.1.	Grafische Verteilungsprüfung der Fähigkeitsschätzer im KFT der Vorstudie.	144
C.2.	Grafische Verteilungsprüfung der Fähigkeitsschätzer im LGVT der Vorstudie.	144
C.3.	Grafische Verteilungsprüfung der Fähigkeitsschätzer im Fachwissen der Vorstudie.	145
C.4.	Schulvergleich der KFT-Fähigkeitsschätzer in der Vorstudie.	146
C.5.	Schulvergleich der LGVT-Fähigkeitsschätzer in der Vorstudie.	146

C.6. Schulvergleich der Fähigkeitsschätzer für das Fachwissen in der Vorstudie.	147
D.1. Regressionslinien für EAP-PVs der Modellaufgaben über den Fähigkeitsschätzern kognitiver Fähigkeiten getrennt nach Schulen.	150
D.2. Regressionslinien für EAP-PVs der Modellaufgaben über den Fähigkeitsschätzern im Fachwissen getrennt nach Schulen.	150
D.3. Regressionslinien für EAP-PVs der Modellaufgaben über den Fähigkeitsschätzern des Leseverstehens getrennt nach Schulen.	151
E.1. Schwierigkeitsverteilung aus der 4-dimensionalen Aufgabenskalierung nach Themen.	154
E.2. Schwierigkeitsverteilung aus der 4-dimensionalen Aufgabenskalierung nach Denkschritten.	154
E.3. Schwierigkeitsverteilung aus der 4-dimensionalen Aufgabenskalierung nach Modellkomponenten.	155
E.4. Schwierigkeitsvergleiche aus der 4-dimensionalen Aufgabenskalierung nach Themen.	155
E.5. Schwierigkeitsvergleiche aus der 4-dimensionalen Aufgabenskalierung nach Denkschritten.	156
E.6. Schwierigkeitsvergleiche aus der 4-dimensionalen Aufgabenskalierung nach Modellkomponenten.	156

Tabellenverzeichnis

2.1. Überblick der ausgewählten theoretischen Rahmungen.	35
2.2. Fachinhaltliche Themen der Studienaufgaben.	39
4.1. Konstruktrelevante Kovariablen.	68
4.2. Überblick des Erhebungsinstruments.	71
4.3. Verwendete R-Pakete.	73
4.4. DETECT-Statistik zur Einschätzung der Multidimensionalität eines Tests.	79
4.5. Orientierungswerte zur Beurteilung von Infit/Outfit. Verändert nach Bond und Fox (2015).	80
4.6. Verwendete Fit-Indices.	81
4.7. Verwendete Informationskriterien.	83
5.1. Umbenennung der Modellkomponenten in Alltagssprache.	86
5.2. Auszählung der Häufigkeiten von den als theoriekonform codierten Zuordnungen durch die SuS.	86
5.3. Anzahl der TeilnehmerInnen sowie Datenmittelwerte der Vorstudie, aufgeschlüsselt nach Schulen, inkl. Datenausschlüssen & MCAR-Quoten.	88
5.4. Vergleich des absoluten Modellfit für 1-PL- & 2-PL-Modell des Fachwissenstests der Vorstudie.	89
5.5. Relativer Modellvergleich zwischen 1-PL- & 2-PL-Modell des Fachwissenstests der Vorstudie.	90
5.6. Relativer Modellvergleich zwischen der einfaktoriellen und der explorativen, zweifaktoriellen 2-PL-Skalierung des Fachwissenstests der Vorstudie.	90
5.7. Personenfähigkeitsschätzungen für die eindimensionale 2-PL-Skalierung des Fachwissenstests der Vorstudie.	91
5.8. Relativer Modellvergleich zwischen 1-PL- & 2-PL-Modell der Modellaufgaben in der Vorstudie.	93
5.9. Überblick der Infit/Outfit-Werte für die 2-PL-Skalierung der Modellaufgaben in der Vorstudie.	93
5.10. Personenfähigkeitsschätzungen für die eindimensionale 2-PL-Skalierung der Modellaufgaben der Vorstudie.	94
5.11. Lineare Regression der über die Plausible Values gemittelten Zusammenhänge der 1-dimensional skalierten Modellnutzungsfähigkeit und Kovariablen über alle Schulen.	95
5.12. Erhebungsrohdaten der Hauptstudie nach Bereinigung, inkl. Datenausschlüssen & MCAR/MAR-Quoten.	98

5.13. Relativer Modellvergleich zwischen 1-PL- & 2-PL-Modell des Fachwissenstests der Hauptstudie.	98
5.14. Personenfähigkeitsschätzung (WLE) für die eindimensionale 2-PL-Skalierung des Fachwissenstests der Hauptstudie.	99
5.15. Überblick der Infit/Outfit-Werte für die eindimensionale 2-PL-Skalierung des Fachwissenstests der Hauptstudie.	99
5.16. Überblick der Infit/Outfit-Werte für die eindimensionale 2-PL-Skalierung der Modellaufgaben in der Hauptstudie.	100
5.17. Relativer Modellvergleich zwischen 1-PL- & 2-PL-Modell der Modellaufgaben.	100
5.18. Relativer Modellvergleich zwischen eindimensionalem 2-PL-Modell mit und ohne Hintergrundvariablen der Modellaufgaben.	101
5.19. Latente Korrelationen der Hintergrundvariablen für die eindimensionale Skalierung.	101
5.20. Relative Modellvergleiche zwischen den verschiedenen Strukturmodellen ohne Hintergrundmodell.	102
5.21. Relative Modellvergleiche zwischen den verschiedenen Strukturmodellen mit Hintergrundmodell und Skalierung mit Within-Item-Dimensionality.	102
5.22. Vergleich der Informationskriterien der verschiedenen Strukturmodellen mit Hintergrundmodell und unter Skalierung mit Within-Item-Dimensionality.	103
5.23. Vergleich des 3- und 4-dimensionalen Theoriemodells mit Zufallsmodellen anhand der Informationskriterien AICc und saBIC.	104
5.24. Latente Korrelationen der Modellnutzungsfähigkeit über die Faktoren der 4-dimensionalen Skalierung.	104
5.25. EAP-Reliabilitäten der Faktoren aus der 4-dimensionalen Skalierung mit Hintergrundmodell.	105
5.26. Deskriptive Darstellung der erreichten Rohpunkte aus den Modellaufgaben nach Testhefttypen.	106
B.1. Schwierigkeitsparameter aus den verschiedenen Skalierungen, inkl. multidim. Aufgabenschwierigkeiten aus der 4-dimensionalen Skalierung.	139
B.1. Schwierigkeitsparameter aus den verschiedenen Skalierungen, inkl. multidim. Aufgabenschwierigkeiten aus der 4-dimensionalen Skalierung (Fortsetzung).	140
B.2. Steigungsparameter bzw. Faktorladungen inkl. multidim. Trennschärfe aus der 4-dimensionalen Skalierung.	141
B.2. Steigungsparameter bzw. Faktorladungen inkl. multidim. Trennschärfe aus der 4-dimensionalen Skalierung (Fortsetzung).	142
C.1. Ergebnisse der Shapiro-Wilk-Tests auf Abweichung von der Normalverteilung in den Fähigkeitsschätzern der Kovariablen.	143
C.2. Ergebnisse der Levene-Tests auf Varianzhomogenität in den Fähigkeitsschätzern der Kovariablen zwischen den Schulen.	143

C.3. Ergebnisse der Kruskal-Wallis-Tests für Schulunterschiede in der Vorstudie. 145

Inhaltsverzeichnis

1. Einführung	15
2. Theoretischer Hintergrund	19
2.1. Erkenntnisgewinnung in den Naturwissenschaften	19
2.1.1. Erkenntnisgewinnung als hypothetisch-deduktives Verfahren	21
2.1.2. Weitere Verfahren für Erkenntnisgewinnung	23
2.2. Modellbegriff	24
2.2.1. Epistemologischer Zugang zu Modellen	24
2.2.2. Modellkomponenten zur chemiespezifischen Modellnutzung	36
2.3. Chemiespezifische Themen für die Modellnutzung	39
2.3.1. Historische Einordnungen	41
2.3.2. Fachliche Beschreibungen	43
2.3.3. Exemplarische Verbindungen der Themen mit modelltheoretischen Aspekten	47
2.3.4. Curriculare Einordnungen	49
2.4. Zusammenfassung	50
2.4.1. Theorieüberblick	50
2.4.2. Zielkonkretisierung	50
2.4.3. Einordnung der Arbeit in das Feld der Kompetenzmodellierung	51
2.4.4. Ableitung des zu prüfenden Konstrukts	52
3. Forschungsfragen & -hypothesen	57
3.1. Forschungsfrage I	57
3.1.1. Hypothese Ia	57
3.1.2. Hypothese Ib	58
3.2. Forschungsfrage II	58
3.2.1. Hypothese IIa	58
4. Methoden	59
4.1. Der Studienplan im Überblick	59
4.2. Die Konstruktion des Erhebungsinstruments	60
4.2.1. Testgütekriterien quantitativer Erhebungen mit multiple-choice Auf- gaben	61
4.2.2. Aufgabenkonstruktion	64
4.2.3. Darstellung und Begründung der Kovariablenauswahl	67
4.2.4. Testheftdesign	69

4.3.	Durchführung der Erhebung	70
4.3.1.	Datenmanagementplan	71
4.3.2.	Umgang mit fehlenden Werten	74
4.4.	Statistische Methoden	75
4.4.1.	Grundlagen der Item Response Theorie	75
4.4.2.	Multidimensionale Item Response Theorie	78
4.4.3.	Modellgeltungsprüfung	79
4.5.	Zusammenfassung des Methodenteils	83
5.	Ergebnisse	85
5.1.	Qualitative Vorarbeiten	85
5.1.1.	Expertenrating	85
5.1.2.	Pilotierung durch offene Aufgaben	86
5.2.	Quantitative Vorarbeiten	87
5.2.1.	Deskriptive Darstellung der Rohdaten	87
5.2.2.	Statistische Skalierungen des Fachwissens- und des Modelltests	89
5.2.3.	Zusammenfassung und Konsequenzen für die Hauptstudie	96
5.3.	Die Hauptstudie	97
5.3.1.	Deskriptive Ergebnisse	97
5.3.2.	Skalierung des Fachwissenstests	97
5.3.3.	Skalierungen der Modellaufgaben	99
5.3.4.	Vergleich der Aufgabenschwierigkeiten	103
5.4.	Zusammenfassung und Beantwortung der Forschungsfragen	107
5.4.1.	Beantwortung der Forschungsfrage I	107
5.4.2.	Beantwortung der Forschungsfrage II	109
6.	Diskussion	111
6.1.	Interpretationen der Befunde	111
6.1.1.	Dimensionalität des Konstrukts	111
6.1.2.	Aufgabenschwierigkeiten	112
6.1.3.	Einfluss der Kovariablen	113
6.1.4.	Rückbindung an die epistemologischen Vorüberlegungen	114
6.1.5.	Vorschläge für unmittelbar mögliche Folgeuntersuchungen	116
6.2.	Zusammenfassung und Ausblick	118
	Literatur	121
A.	Abkürzungsverzeichnis	137
B.	Aufgabenparameter	139
B.1.	Schwierigkeitsparameter	139
B.2.	Steigungsparameter	141
C.	Statistische Prüfungen der Kovariablen für die Vorstudie	143

D. Zusammenhänge der Kovariablen mit den PVs der Modellaufgaben in der Vorstudie	149
E. Schwierigkeitsvergleiche der Aufgaben in der Hauptstudie	153
F. Konstruktionsmanual	157
G. Testheftdesign	163
H. Beispielaufgaben	165
I. Lebenslauf inkl. Publikationen	169
J. Eigenständigkeitserklärung	175
K. Danksagung	177

1. Einführung

Die Konstruktion und Nutzung von Modellen für naturwissenschaftliche Erkenntnisgewinnung sind mächtige Werkzeuge menschlicher Kognition, denn sie ermöglichen die Erschließung der Welt außerhalb sinnlicher Wahrnehmung. Sie sind historisch mit außergewöhnlichen Persönlichkeiten wie bspw. Demokrit, Descartes und Dalton verbunden (vgl. Weyer, 2018). Sie kanonisieren und harmonisieren die Beschreibung, Erklärung sowie die Vorhersage naturwissenschaftlicher Phänomene. Ihre umfangreiche Bedeutung ist aber nicht nur in historischer Weise evident. Sie werden für den Umgang mit zukünftigen Herausforderungen in einer globalisierten Wissensgesellschaft als tragende Säule für Bildung antizipiert (vgl. Harlen, 2015; Harlen & Bell, 2010; National Research Council, 2012; OECD, 2019a). Weil außerdem berufliche Tätigkeiten mit geringer Qualifizierung, d. h. für die ausschließlich kognitive Fähigkeiten ohne großes Transferpotential notwendig sind, rückläufig sind (European Centre for the Development of Vocational Training, 2019), ist das Erlernen der Modellnutzung in Bildungsstandards festgeschrieben. Beispielsweise helfen Modelle laut dem Rahmenlehrplan des Fachs *Naturwissenschaften* der 5./6. Jahrgangsstufe des Landes Berlin „[...] sich eine Vorstellung von der Wirklichkeit zu machen, die der Anschauung nicht zugänglich ist.“ (Landesinstitut für Schule und Medien Berlin-Brandenburg, 2017, S. 7). Darüber hinaus verwenden Lernende laut des Rahmenlehrplans Chemie

„[...] ein Modell als eine idealisierte oder generalisierte Darstellung eines existierenden oder gedachten Objektes, Systems oder Prozesses. Die Auswahl eines geeigneten Modells unter Beachtung der Fragestellung und das kritische Reflektieren des Modells sind bedeutsamer Bestandteil der naturwissenschaftlichen Erkenntnisgewinnung.“ (Landesinstitut für Schule und Medien Berlin-Brandenburg, 2016, S. 6)

Diese normativen Vorgaben sind an der begrifflichen Oberfläche plausibel. Betrachtet man allerdings überblicksweise die Ergebnisse aus der Forschung zu Modellen und Modellierung in kompetenzorientierten Lehr-/Lernkontexten (Gilbert & Justi, 2016, S. 29 ff.), fallen inkonsistente und lückenhafte theoretische Beschreibungen und entsprechende empirische Konsequenzen auf. In der Folge kann einerseits eine schwache Forschungslage (Louca & Zacharia, 2012, S. 481 ff.), sowie ein „[...] need for a unifying framework [...]“ formuliert werden (Nicolaou & Constantinou, 2014, S. 71).

Eine basale Annahme der vorliegenden Arbeit ist es, dass in der fachdidaktischen Forschung nicht ausreichend expliziert wird, welche erkenntnistheoretischen Grundannahmen bei der Untersuchung von Modellen gemacht werden. Krüger et al. (2018) stellen beispielsweise umfänglich dar, welche Ansätze für das Modellverständnis im Feld

1. Einführung

genutzt werden. Der gemeinsame Nenner sei dabei die Subjekt-, Zweck- und Zeitbezogenheit. Sie identifizieren aber ausschließlich empirische Forschungsdesiderate (Krüger et al., 2018, S. 152) und rekurren für das Erstaunen (ebd., S. 141) über die uneinheitlichen Begrifflichkeiten lediglich auf die immerwährende Notwendigkeit für die Prüfung und Revision von Modellen. Es scheint also geboten, erkenntnistheoretische Transparenz herzustellen. Wenn das im Allgemeinen funktionieren würde, dann käme darüber hinaus noch ein fachspezifisches Moment hinzu. Der Modellbegriff mag verallgemeinerbar sein, die konkrete Modellnutzung und deren empirische Struktur sind aber nicht unmittelbar von einem Fach auf ein anderes übertragbar. Ein deutliches Beispiel dafür ist die Dimension *Vergleich mit dem Original* aus dem *Kompetenzmodell der Modellkompetenz* (Upmeyer zu Belzen & Krüger, 2010), die bei der Nutzung von Atommodellen überhaupt keine empirische Kategorie sein kann: Niemand kann ein Atommodell mit seinem Original vergleichen. Ein weiteres Beispiel ist das Lotka-Volterra-Modell von Räuber-Beute-Beziehungen, das auf ökonomische Zusammenhänge übertragen werden kann und Prädiktionskraft hat. Es ist aber bei diesem Transfer enorm auf seine Prämissen angewiesen,¹ sodass es mindestens nicht überraschen sollte, wenn die entsprechenden Vorhersagen gelegentlich kollabieren. Die Explikation der jeweiligen Vorannahmen sind also einerseits enorm wichtig für Forschung zur Modellnutzung. Andererseits wird dem selten umfänglich Rechnung getragen. Schwarz et al. (2009) verwenden beispielsweise eine Modelldefinition, die sich als Schnittmenge aus empirischen Studien verstehen lässt und die Modelle als vereinfachte Repräsentationen identifiziert (vgl. hierzu auch Giere, 2004). Krell et al. (2015) beziehen sich auf Schwarz et al. (2009) und finden beim Fächervergleich zwischen Biologie, Chemie und Physik eine Abhängigkeit des epistemischen Verständnisses vom Konstrukt des *meta-modeling knowledge* (vgl. bspw. Schwarz & White, 2005). Obwohl Gilbert und Justi (2016) ein differenziertes Bild von Lernsteigerungsnachweisen präsentieren (ebd., S. 90 ff.), attestieren sie dem Feld, dass *kein* expliziter Unterricht *über* Modelle untersucht wurde. Setzt man mit Grosslight et al. (1991) eine historische Klammer um den Befund über die Notwendigkeit von expliziten Lerngelegenheiten mit und über Modelle bei SuS, kann mindestens das chemiespezifische Defizit in der Begriffsklärung und den empirischen Erkenntnissen festgestellt werden. Das vorliegende Projekt versucht exakt dieses Problem zu bearbeiten und eine durchgehende Linie zu zeichnen, die die epistemologischen Prämissen und Konklusionen aus dem Theorieteil konsequent in seinen empirischen Teil einarbeitet und in der Diskussion explizit berücksichtigt.

Zusammengefasst gilt es folgende Analysen vorzunehmen: Welche epistemologischen Basistheorien tragen zu einer Begriffsklärung bei? Zwischen der Frage *Was ist ein Modell?* und *Welchem Zweck dient ein Modell?* ist dabei zu unterscheiden und in der Folge Modellnutzung und Modellierung differenziert darzustellen. Der Einflussgrad der Subjektivität eines Modells auf seine Generalisierbarkeit ist zu untersuchen, d. h. es wird zwischen kanonisch-historisch bewährten und momentgebundenen ad-hoc Modellen zu unterscheiden sein. Die Grenzen der Fachspezifität von Modellen werden abgesteckt

¹Dazu gehört die Annahme eines *homo oeconomicus*, also eines Menschen, dessen Intentionen von der Nutzenmaximierung dominiert werden und der sich streng rational verhält.

und eine chemiespezifische Interpretation angeboten, sowie eine Bewertung des dargelegten Theorierahmens vorgenommen. Daraus ergeben sich Schlussfolgerungen, die zur kohärenten Konstruktion von chemiespezifischen Aufgaben führen, um anschließend zu testen, ob diese Aufgaben a) überhaupt von Schülerinnen und Schülern gelöst und b) die theoretisch angenommenen Teildimensionen der Aufgaben empirisch wiedergefunden werden können.

2. Theoretischer Hintergrund

Der Theorieteil der vorliegenden Arbeit ist in vier Abschnitte unterteilt. In Abschnitt 2.1 wird der Begriff der Erkenntnisgewinnung definiert. Zum Zweck der empirischen Erhebung wird diese Definition eng gefasst sein. In Abschnitt 2.2 geht es um die Darstellung des verwendeten Modellbegriffs. Dieser wird aus einer erkenntnistheoretischen Perspektive aufgearbeitet. Dabei wird das Nutzen von Modellen zum Durchlaufen des hypothetisch-deduktiven Verfahrens im Vordergrund stehen und über vier Modellkomponenten einer empirischen Untersuchung zugänglich gemacht werden. Weil Erkenntnisgewinnungsprozesse einen inhaltlichen Bezugsrahmen benötigen (vgl. Subjekt- und Zweckgebundenheit von Modellen in Abschnitt 2.2 nach Mahr, 2008), wurden für die Aufgaben fünf Themen identifiziert, deren fachliche, historische und curriculare Zusammenhänge im dritten Teil dargestellt werden (Abschnitt 2.3). Abschnitt 2.4 führt die drei vorangegangenen zusammen, expliziert in der Folge die Forschungslücke und präsentiert das theoretische Konstrukt, das der empirischen Prüfung zugeführt werden soll.

2.1. Erkenntnisgewinnung in den Naturwissenschaften

Der Begriff Erkenntnisgewinnung bedarf für eine empirische Untersuchung einer Definition, weil er mit vielen Bedeutungen belegt werden kann. Wissenschaftliche Erkenntnis könne nach Popper (1935) allein durch falsifizierbare Hypothesen das subjektunabhängige Projekt vom Verstehen der Welt bedienen. Die berühmte Illustration des sogenannten *Induktionsproblems* durch die Beobachtung von Schwänen¹ ist dabei nur der Auftakt für eine umfassende Theorie empirischer Erkenntnisgewinnung. Eine tiefere Analyse der *Logik der Forschung* (Popper, 1935) ist für die vorliegende Studie nicht zielführend. Das *Induktionsproblem* spielt aber an der Grenzfläche von Modelltheorie und Erkenntnisgewinnung als Kontrast zum hypothetisch-deduktiven Verfahren (vgl. Abschnitt 2.1.1) eine wichtige Rolle. Das gilt sowohl für die normativen Rahmenbedingungen, die theoretische Aufarbeitung, als auch für die praktische Operationalisierung der Testaufgaben dieser Arbeit.

Erkenntnisgewinnung kann im konkreteren Zusammenhang von Lehr-/Lernumgebungen in verschiedener Weise gefasst werden, bspw. als vorläufiger Endpunkt von Wissensakku-

¹Wer in seinem Leben nur weiße Schwäne gesehen hat, kann kein verallgemeinerbares Argument bereitstellen um zu behaupten, dass *alle* Schwäne auf der Welt weiß sind. Es ist lediglich unter Vorbehalt möglich, eine subjektive Bestandsaufnahme zu formulieren: Die Beobachtung aller Schwäne bisher, kann zur Annahme führen, dass alle Schwäne weiß sind. Sobald der erste schwarze Schwan beobachtet wird, muss die Annahme zurückgewiesen werden.

2. Theoretischer Hintergrund

mulation. Dazu gehört das Konzept des Spiralcurriculums und der Wissensvernetzung, bspw. beschrieben bei Bernholt et al. (2018). Für eine chemiespezifische Interpretation sehen Erduran und Kaya (2019, S. 36 ff.) naturwissenschaftliche Theorien, Gesetze oder Modelle als unterschiedliche Ausprägungen von Wissen als Grundlage eines sich ständig entwickelnden Erkenntnisfortschritts.² Auch der Prozess des Konzeptwechsels (Vosniadou, 2012) kann als Erkenntnisgewinnung in Lehr-/Lernkontexten verstanden werden. Denn in diesem Sinne soll, unter der Annahme Naturwissenschaften seien semantische Ordnungssysteme, ein sukzessiv kohärenterer Umgang mit Phänomenen erzielt werden. Zwei weitere Ansätze bestehen darin, Erkenntnisgewinnung als systematische Reduktion von Unsicherheit (Kirch, 2012), oder als empirischen Problemlöseprozess (Dunbar & Klahr, 2012; Gut-Glanzmann & Mayer, 2018) zu verstehen. Diese Facetten schließen sich nicht notwendigerweise gegenseitig aus, sie zielen aber auf eine derart große Bandbreite von Konstrukten aus Lern- und Wissenschaftstheorien ab, dass eine empirische Fassung und Interpretation im Rahmen der vorliegenden Arbeit nicht zu leisten wäre. Für die vorliegende Studie wird der Erkenntnisgewinnungsprozess im Anschluss an Nehring et al. (2016) als Dreischritt operationalisiert (vgl. auch Vorholzer et al., 2016). Zunächst ist der Lerner mit einem Phänomen, bspw. einer chemischen Reaktion konfrontiert. Diese soll systematisiert, erklärt oder reproduziert werden, ggf. auch ohne, dass Kenntnisse über die chemisch-physikalischen Zusammenhänge vorliegen. Auf Grundlage eines Phänomens soll eine Frage gestellt werden, die in eine Hypothese überführt wird (Schritt 1). Es folgt die Entwicklung und Durchführung eines geeigneten Untersuchungsverfahrens, in dem hypothesenbezogene Daten (Zählungen, Messungen, Kategorisierungen) gesammelt werden (Schritt 2). Am Ende steht der systematische Rückbezug dieser Daten auf die Hypothese, sodass diese zurückgewiesen oder beibehalten werden kann (Schritt 3). Dieses hypothetisch-deduktive Verfahren bedarf im Folgenden einer Erläuterung, die in Abschnitt 2.1.1 gegeben wird.

Dieser Referenzrahmen steht nicht ohne Kritik in den Debatten um Lehr-/Lernprozesse (Rönnebeck et al., 2016), er steht aber in kohärenter Beziehung zu etablierten theoretischen Ausarbeitungen (Andersen & Garcia-Mila, 2017; Lehrer & Schauble, 2003), empirischen Ergebnissen im Feld der Chemiedidaktik (Nehring et al., 2015; Scherer, 2014), oder auch interpretiert als naturwissenschaftlicher Diskurs im Biologieunterricht (Ergazaki & Zogza, 2005). Pedaste et al. (2015) geben einen Überblick zu Konzeptualisierungen dieses Ansatzes, die sich alle auf dieser Linie bewegen. Insofern ist die gewählte Perspektive auf Erkenntnisgewinnung sowohl anschlussfähig zu etablierten Theorien in der empirischen Bildungsforschung, als auch offen für Erweiterungen und Diskussionen ihrer Grundannahmen und Konsequenzen (vgl. Abschnitte 2.4.3 und 6.1.5).

²In diesem Zusammenhang wird hier auf die Diskussion darüber hingewiesen, ob *Wissen* überhaupt eine sinnvolle erkenntnistheoretische Kategorie sein kann (Beckermann, 2001). Siehe dazu auch die sog. *Gettier-Fälle* (Gettier, 1963).

2.1.1. Erkenntnisgewinnung als hypothetisch-deduktives Verfahren

Als erster namensgebender Teilbegriff wird eine Hypothese hier als Aussage verstanden, die durch ihre formallogische Struktur (vgl. *Regeln* in Abschnitt 2.2.2) einen empirisch prüfbareren Zusammenhang von Phänomenen formuliert und grundsätzlich falsifizierbar gehalten sein muss (Popper, 1935). Sie kann unterschiedliche Formen annehmen und grundsätzlich sowohl der Beginn, als auch das Ergebnis einer naturwissenschaftlichen Untersuchung sein: Wer sich für Schwäne interessiert kann aus Beobachtungsdaten die Hypothese ableiten, dass diese Tiere ausschließlich weißes Gefieder besitzen (vgl. Abschnitt 2.1.2). Ebenfalls möglich ist es, die Aussage nach der Beobachtung des ersten schwarzen Schwans zurückzuweisen. Die Hypothese leitete in der Folge die gesamte Untersuchung und ihre Falsifizierung stellt den Endpunkt des Verfahrens dar. Die hypothesenbezogene Arbeit im Chemieunterricht ist – insbesondere bezüglich der Modellnutzung – nachweisbar unterrepräsentiert (Nehring et al., 2015) und für Erkenntnisprozesse im Schulunterricht ist auch kein systematischer Rahmen erkennbar. Die Begründung dafür liegt, erstens, sowohl in der empirischen Unsicherheit über das exakte Verhältnis zwischen der Struktur der Hypothesennutzung und der Aufgabenschwierigkeit für SuS im naturwissenschaftlichen Unterricht (Krell, 2018), als auch in einem Mangel an begrifflicher Konkretisierung (Vorholzer et al., 2016). Zweitens sind die Verknüpfungen von formalen Schlüssen und Modellen noch nicht systematisch theoretisiert (Gilbert & Justi, 2016, S. 97 ff.), bzw. als Forschungsdesiderata in der Modell- und Argumentationsforschung identifizierbar (Nitz & Fechner, 2018; von Aufschnaiter & Prechtel, 2018). Osborne und Patterson (2011) diskutieren, dass eine nicht sorgfältig vorgenommene Unterscheidung zwischen einem hypothesenbezogenem Argument und einer Erklärung als Hypothese, als zwei aufeinander bezogene Operationen auch in unterrichtspraktischer Hinsicht ein relevantes Hindernis für Lehrende und Lernende ist.

Der zweite namensgebende Teilbegriff – die Deduktion – rekonstruiert das naturwissenschaftliche Vorgehen als eines von zahlreichen Argumentationsmustern (Salmon, 2003). Es schließt aus Obersätzen bzw. Prämissen mithilfe eines formallogischen Schlussprinzips von allgemein gültigen Regeln auf Einzelfälle (Tetens, 2010). Das Prinzip stammt in seiner systematischen Aufarbeitung ursprünglich von Aristoteles, ist auch Bestandteil moderner erkenntnistheoretischer Untersuchungen (Bara & Bucciarelli, 2000; Byrne & Tasso, 1999; Johnson-Laird, 1999) und findet sich als Ausgangspunkt für empirische psychologische Untersuchungen menschlicher Kognition (Johnson-Laird et al., 1992; Khemlani et al., 2013). Ein beliebtes Beispiel zur Illustration lautet:³

- *Prämisse 1*: Alle Menschen sind sterblich.

³Im ersten Beispiel wird mit dem Prinzip der Transitivität als Schlussprinzip gearbeitet, obwohl auch der sogenannte *modus tollens* verwendet werden kann. Auf diese strenge Formalisierung insbesondere der Schlussprinzipien wird hier verzichtet um den Lesefluss aufrechtzuerhalten, zumal die Konklusionen aus verschiedenen Rekonstruktionsstrategien abgeleitet werden können. In der entsprechenden Literatur häufig verwendete Notationen werden in Abschnitt 2.2.2 gezeigt. Die Semantik geht durch die Darstellung in natürlicher Schriftsprache nicht verloren.

2. Theoretischer Hintergrund

- *Prämisse 2*: Aristoteles ist ein Mensch.
- *Schlussprinzip*, (Transitivität): Wenn alle x die Eigenschaft E besitzen und y Teilmenge von x ist, dann besitzt y die Eigenschaft E .
- *Konklusion*: Aristoteles ist sterblich.

Für die Anwendung auf den Chemieunterricht wird das folgende Beispiel angeführt:

- *Prämisse 1*: Ein Stoff brennt, wenn ausreichend Sauerstoff an seine Oberfläche gelangt und die Zündenergieschwelle überschritten wird.
- *Prämisse 2*: Es ist ausreichend Sauerstoff an der Oberfläche des Stoffs.
- *Prämisse 3*: Die Zündenergieschwelle wird überschritten.
- *Schlussprinzip* (sog. modus ponens): Wenn x und y zutreffen, dann z und x und y treffen zu.
- *Konklusion*: Der Stoff brennt.

Weil das Zutreffen der Prämissen 2 und 3 notwendige Bedingungen für das Zutreffen der Konklusion sind, könnte das Nichtbrennen des Stoffs in weiteren Schemata auf das Nichtzutreffen seiner notwendigen Bedingungen zurückgeführt werden. Diese Zusammenhänge erscheinen in Sachverhalten mit niedriger Komplexität trivial. Ihre Explikation und Bedeutung für ein empirisches Forschungsvorhaben wird aber als notwendig erachtet, weil sie in zahlreichen theoretischen Ansätzen (nur) implizit enthalten sind. Das wird folgend an zwei Beispielen aus zusammenfassenden Arbeiten illustriert.

Stephens und Clement (2012) definieren Gedankenexperimente in naturwissenschaftlichen Lehr-/Lernumgebungen, indem sie ihnen eine systembeschreibende Funktion zuweisen: Sie seien „[...] the act of considering an untested, concrete system designed to help evaluate a scientific concept, model, or theory – and attempting to predict aspects of the system’s behaviour.“ (ebd., S. 160). Dieser Prozess kann nur als hypothetisches Verfahren funktionieren, wenn dessen Vorhersagen an Beobachtungen überprüft und mit diesen per oben beschriebener Argumentationsketten sprachlich verknüpft werden. Zoller und Nahum (2012) unterscheiden zwischen *lower-order cognitive skills* (LOCS) und *higher-order cognitive skills* (HOCS) und ordnen letzterem bspw. problemlösendes Denken, sowie die Fähigkeit zum Transfer zu (ebd. S. 213 f.). Sie beschreiben den von ihnen als notwendig erachteten Übergang von LOCS zu HOCS als leistbar mit einer Reihe von Aktivitäten im Klassenraum (ebd., S. 217 ff., sinngemäße Begriffsübersetzungen durch Autor), ohne aber zu konkretisieren, wie die Betonung von *forschend-entwickelndem Unterrichtsgespräch* (vgl. auch Streller et al., 2019) *selbstgesteuertem Lernen*, oder *Selbstreflexion* operationalisiert werden könnte. Setzte man das Einüben hypothesenorientierten Schlussfolgerns in die angegebene Liste, so würden einige Punkte darin aufgehen. Die Förderung der *Argumentationsfähigkeit* und die *Transferfähigkeit* würden dann verallgemeinert, weil formale Begriffsverbindungen wie *wenn-dann* und *entweder-oder* kontext- und fachunabhängig sind. Gleichzeitig können sie in beliebigen Fachkontexten ausprobiert

und angewendet werden. Damit würden dann ggf. gewünschte Fachinhalte aufrecht erhalten, ohne auf der Ebene der LOCS stehenbleiben zu müssen.

Diese Form der Prüfung von Aussagen, d. h. deduktives Schlussfolgern unter der Annahme der Wahrheit von Prämissen und die empirische Prüfung des Zutreffens von Konklusionen, kann in der Folge mit einem hypothesengeleiteten Prüfprozess in den Naturwissenschaften identifiziert werden. Ist die Konklusion eine empirisch beobachtbare Aussage, so muss bei ihrem Nichtzutreffen entweder eine notwendige Bedingung fehlen, oder mindestens eine der Prämissen falsch sein. In jedem Fall können die notwendigen und hinreichenden Bedingungen eines Sachverhalts auf diese Weise ausgelotet und ggf. eingegrenzt werden. Dies wird in der vorliegenden Arbeit als hypothetisch-deduktive Erkenntnisgewinnung bezeichnet.

2.1.2. Weitere Verfahren für Erkenntnisgewinnung

Für die Formalisierung von Schlüssen werden in der Literatur üblicherweise drei Grundstrukturen angegeben. Dabei handelt es sich um die *Deduktion* (siehe Abschnitt 2.1.1, sowie die *Induktion* und die *Abduktion* (Aliseda, 2000; Flach, 2000). Üblicherweise wird die Induktion als Gegenstück zur Deduktion dargestellt (Bara & Bucciarelli, 2000). Während bei letzterer von der allgemeinen Regel auf den Einzelfall geschlossen wird, formuliert erstere eine allgemeine Regel auf Basis von Einzelfällen. Aus den Arbeiten des Logikers Charles Sanders Peirce ist, besonders im Zusammenhang mit Hypothesen, der Begriff der Abduktion eingeführt worden, der häufig als *Inferenz zur besten Erklärung* formuliert wird (Aliseda, 2000). Diese Schlussform erklärt – für formallogische Veröffentlichungen ungewöhnlich diffus mit Begriffen wie „überraschend“ (ebd., S. 48) – den Übergang von vorliegenden Beobachtungen zu erkenntniserweiternden Aussagen. Die Terminologie und die Diskussion haben durchaus Einfluss auf naturwissenschaftliche Erkenntnisprozesse, weswegen sie hier für den Begriffsüberblick erwähnt werden. Ihre konkreten Zusammenhänge können hier aber nicht von ihren ersten Prinzipien abgeleitet und dargestellt werden. Es wird lediglich zusammengefasst, um das Arbeiten mit Hypothesen in naturwissenschaftsdidaktischer Hinsicht eine umfänglichere Rahmung zu geben. Aliseda (2000) fasst für Schlussfolgerungen mit Anspruch auf Allgemeingültigkeit die Induktion als einen Spezialfall der Abduktion auf (ebd., S. 47) und abstrahiert beides zum *erklärenden Schlussfolgern* (vgl. „explanatory reasoning“, Osborne & Patterson, 2011). Weil die Abduktion aus sich heraus kein Kriterium für eine *beste* Erklärung liefert, ist sie trotz ihrer Plausibilität umstritten und läuft bei der Generalisierung stets Gefahr, die Schlussfolgerung immer schon vorauszusetzen (vgl. Flach und Kakas (2000) und Paavola (2004), sowie das *Erfolgsargument* des wissenschaftlichen Realismus in Abschnitt 2.2.1).

Zusammenfassung

Erkenntnisgewinnung wird hier als Ziehen von Schlüssen definiert und auf formallogische Operationen zurückgeführt. Die wichtigsten Typen von Schlussfolgerungen sind die Induktion und die Deduktion, wobei letztere, als Schluss von allgemeinen Regeln auf spezielle Sachverhalte, im Fokus steht. Dies begründet sich einerseits im sachlogischen

2. Theoretischer Hintergrund

Zusammenhang vom Verständnis der Modellnutzung (s. u.), sowie aus dem Anschluss an normative Vorgaben durch die Bildungsstandards, die das Nutzen von Modellen als explizit deduktives Verfahren als zu erwerbende Kompetenz festschreiben. Die vorliegende Arbeit fokussiert auf dieses hypothetisch-deduktive Vorgehen, indem diese spezifischen Art und Weise mit naturwissenschaftlichen Sachverhalten umzugehen mit den *Modellkomponenten* in Abschnitt 2.2.2 verknüpft wird. Das spiegelt sich in der Folge in der empirischen Messbarmachung wider (vgl. Abschnitt 2.4.4).

2.2. Modellbegriff

Den allgemeinen Überlegungen zur naturwissenschaftlichen Erkenntnisgewinnung im Chemieunterricht folgt in diesem Abschnitt die Nutzung von Modellen zu eben jenem Zweck. Dazu wird ein erkenntnistheoretischer Modellbegriff entwickelt, der flexibel genug ist, um verallgemeinerbar auf beliebige Kontexte (vgl. Abschnitt 2.3) angewendet werden zu können. Gleichzeitig soll er für die fachspezifischen Anforderungen der Chemie, d. h. für den Bezug auf nicht wahrnehmbare, kleinste Teilchen, geeignet sein und mithilfe der vier *Modellkomponenten* (Abschnitt 2.2.2) einen Übergang zwischen den theoretischen Überlegungen hin zu einer quantitativen Fragebogenstudie gewährleisten.

2.2.1. Epistemologischer Zugang zu Modellen

Für die Modelldefinition dieser Arbeit, ist der Rückgriff auf erkenntnistheoretische Arbeiten notwendig, die zunächst nicht explizit an naturwissenschaftliche Überlegungen angebunden sind. Die dargestellte, epistemologische Diskussion nimmt durchaus Bezug auf naturwissenschaftliche Sachverhalte:

„Attention to the epistemological side of model construction, functioning and testing is of special current importance because of the unprecedented importance being given to *Nature of Science* objectives in science curricula around the world.“ (Matthews, 2007, S. 647)

Der Fokus liegt aber – gemäß der typischen Vorgehensweise in der Erkenntnistheorie – auf den *prinzipiellen* Zusammenhängen von Sachverhalten, die auch ohne empirische Entsprechungen eingesehen werden können und sollen. Dieser Exkurs ist von besonderer Bedeutung um den Zusammenhang zwischen einem hypothetisch-deduktiven Vorgehen (vgl. Abschnitt 2.1.1) und den normativen Vorgaben seitens der Bildungsstandards plausibel zu machen. Diese kommen nämlich nicht ohne bestimmte, erkenntnistheoretische Setzungen aus. Insofern ist es notwendig, jene herauszuarbeiten und ihren spezifischen Beitrag zum empirischen Teil der vorliegenden Arbeit (d. h. Strukturbestimmung & Messbarmachung der Modellnutzung für hypothetisch-deduktives Denken im Chemieunterricht) präzise zu benennen. Die Fassung des Modellbegriffs in erkenntnistheoretischer Hinsicht erfolgt über die Theorien des *wissenschaftlichen Realismus* (Devitt, 2006), des *konstruktiven Empirismus* (van Fraassen, 1980), der *Allgemeinen Modelltheorie* (Stachowiak, 1973), sowie einer Synthese der Theorien über den Ansatz von Modellen als *epistemische Werkzeuge* (Knuuttila, 2005), die die Zweckbindung in den Vordergrund stellt.

Diese Zweckbindung wird vorher im *Modell des Modellseins* nach Mahr (2008) aufgearbeitet. Modelle werden im Anschluss an die genannten Ausführungen als zweck- und momentgebundene Konstrukte verstanden, für den empirischen Teil (Kapitel 4 und 5) als solche operationalisiert und im Diskussionsteil (Kapitel 6) entsprechend reflektiert werden.

Wissenschaftlicher Realismus

Der wissenschaftliche Realismus formuliert eine Theorie, die einem intuitiven Zugang auf Modelle folgt. Von Devitt (2006) wird der Leitsatz formuliert, der im Rahmen dieser Dissertation maßgeblich ist:

„For example, why are all the observations we make just the sort we would make if there were atoms? Answer: because there *are* atoms. Sometimes the realist goes further: it would be 'a miracle' that theories were so successful if they weren't approximately true. Realism does not just have the best explanation of success, it has *the only good* explanation.“ (ders., S. 5)

Dieses Argument wird als das *Erfolgsargument* bezeichnet. Naturwissenschaftliche Vorhersagen, die aus Beobachtungen gefolgert wurden, sind nachweislich immer präziser geworden. Es ist schwer zu bestreiten, dass die Hypothese der Existenz kleinster Teilchen die Welt exakter erklärt und vorhersagt, als die Hypothese eines Materiekontinuums (Barke et al., 2015, S. 8). In dieser Hinsicht sind Modelle beschreibende und prädiktive Hilfsmittel, deren benannten Bestandteilen (bspw. Atome mit Eigenschaften wie Masse oder räumlicher Ausdehnung) ein Wahrheitsgehalt zukommt. Ihre Existenz trifft zu, weil sie die beste Erklärung darstellen.

Es sei an dieser Stelle noch darauf verwiesen, dass graduelle Wahrheitszuweisungen („[...] approximately true [...]“, s. o.) durchaus umstritten sind. Für eine ausführliche Diskussion dazu und die Schwierigkeiten des Erfolgsarguments vgl. Mizrahi (2011). Für diese Arbeit steht das Erfolgsargument in seiner allgemeinen Form im Vordergrund, weil die anderen Argumente für (Abduktion & methodologischer Erfolg) oder gegen den wissenschaftlichen Realismus formal daraus abgeleitet werden können (vgl. Devitt, 2006, S. 5 ff.).

Konstruktiver Empirismus

Das Theoriegebäude des *konstruktiven Empirismus* wendet sich gegen die Wahrheitsannahme von Sachverhalten (hier: Komponenten von Modellen für hypothetisch-deduktives Denken):

„After deciding that the language of science must be literally understood, we can still say that there is no need to believe good theories to be true, nor to believe *ipso facto* that the entities they postulate are real.“ (van Fraassen, 1980, S. 11 f.)

2. Theoretischer Hintergrund

Van Fraassen (1980) führt aus, dass der Wahrheitsgehalt (natur-)wissenschaftlicher Theorien und ihrer Subsysteme (bspw. Modelle und/oder Definitionen) überhaupt keine notwendige Prämisse ist. Der zentrale Bezugspunkt ist es „[...] das Phänomen zu retten [...]“ (eigene Übersetzung, van Fraassen, 1980, S. 12). Solange eine Theorie empirisch adäquat sei, benötige sie als Akzeptanzkriterium allein die Überzeugung, dass ihre Angemessenheit zutrifft. Es ist folglich unerheblich, ob Atome tatsächlich existieren. Weil die Beschreibung der Welt hervorragend funktioniert, wenn angenommen wird, dass sie aus Atomen aufgebaut ist, ist diese Annahme empirisch adäquat und insofern hat die Wissenschaft eine zielführende Theorie formuliert. Aus dem Zusammenspiel von Beobachtung und dem Überzeugt-Sein des Beobachtenden, der die Verbindung von Theorie und Phänomen konstruiert, ergibt sich dann gerade die Bezeichnung des *konstruktiven Empirismus*.

Problematisch an dieser Einordnung ist, dass der *konstruktive Empirismus* nicht darstellen kann, was seine Bezugspunkte sind, wenn über konkrete Sachverhalte gesprochen wird. Während der Realist kein Problem damit hat ein Atom zu postulieren und dieses Postulat dann mit empirischen Daten zu unterlegen, muss der konstruktive Empirist im strengen Sinne damit leben, dass all seine Erkenntnisse und Annahmen hypothetisch bleiben. Das ist zumindest für den wissenschaftlichen Alltag aber problematisch. Niemand würde auf einem Chemiekongress die Existenz von Molekülen – in der Art in der sie beispielsweise als Strukturformeln dargestellt werden – in Frage stellen. Ohne einen (zumindest vorläufigen) Wahrheitsanspruch, kann die empirische Angemessenheit eines Sachverhalts überhaupt nicht untersucht werden.

Während diese abstrakte Darstellung zunächst weit entfernt von Chemieunterricht zu sein scheint, hat das dargestellte, konstruktive Moment ganz konkrete Auswirkungen auf Lern- und Bildungstheorien. Konstruktivistische Theorien, die vom Subjekt und seiner geistigen Konstruktion der Welt ausgehen, dominieren den Diskurs, sowohl in der empirischen und theoretischen Bildungsforschung (Riemeier, 2007), als auch im alltäglichen Schulunterricht (Streller et al., 2019, S. 89 ff.). Zwar führt allein die Quantität und Präsenz einer Theoriefamilie keine Notwendigkeit mit sich, plausibel ist die Stoßrichtung des modernen Konstruktivismus aber mindestens, wenn eine Unterscheidung zwischen dem beschriebenen Sachverhalt und der beschreibenden Theorie getroffen wird. Während submikroskopische Partikel unabhängig von einem Beobachter existieren mögen, tun Modelle zu diesen Partikeln dies *nicht*. Nicht nur sind verschiedene Beschreibungs- und Vorhersagerahmen für die Welt des Kleinsten konzipiert worden (bspw. Bohr'sches Atommodell, Welle-Teilchen-Dualismus, Lewis-Strukturformeln für Moleküle) – sie existieren nebeneinander in Lehrbüchern und widersprechen sich in wesentlichen Grundannahmen gegenseitig. Während das möglicherweise als Streit zwischen verschiedenen Wissenschaftsschulen noch akzeptabel wäre, muss aber außerdem festgestellt werden, dass ein und dieselbe Person die Annahme fest definierter Kreisbahnen von Elektronen um einen Atomkern und feste Bindungen mit lokalisierbaren Ladungsdichten zwischen Atomkernen nicht zur Deckung bringen kann. Zu solch einer „Inkommensurabilität“ von Modellvorstellungen *innerhalb* einer Person tragen bspw. Flores-Camacho et al. (2007) bei. Dieses Problem trifft exakt auf Chemieunterricht zu, weil beispielsweise in Schulbüchern ständig zwischen verschiedenen, sich teils widersprechenden Modelldarstellungen

gewechselt wird. Dazu gehören die Wechsel verschiedener Repräsentationen chemischer Strukturen (Arnold, 2018, S. 68 ff.), der Sprung vom Kern-Hülle-Modell zu fixen Elektronenkreisbahnen (Bohrmann-Linde et al., 2018, S. 110 f.), sowie die Darstellung eines Lösungsvorgangs mit dem Kugelteilchenmodell, ohne auf zwischenmolekulare Wechselwirkungen einzugehen (Arnold & Dietrich, 2010, S. 43). Das hängt mit dem genuin hypothetischen Charakter von Modellen zusammen, der auch wissenschaftshistorisch gedeutet werden kann. Modelle wurden und werden stets im Modus des *Was-Wäre-Wenn* verwendet (vgl. Kind & Osborne, 2017, S. 12 ff.). Selbst wenn die Modellierung eines Sachverhalts exakte Vorhersagen liefert, steht sie unter der Bedingung bestimmter Annahmen. Dazu gehört in der Chemie beispielsweise die Annahme bestimmter Teilcheneigenschaften. Diese werden allerdings nicht jedes mal expliziert, weil sich ihre Prädiktionskraft für konkrete Anwendungen bewährt hat. Keine organische Chemikerin formuliert in den Aufzeichnungen für ihre Kolleginnen, dass eine ihrer Grundannahmen die Beschreibung von Atomen als Kugelteilchen und die von Elektronenbindungen als bewegliche kleine Stäbchen ist. Sie würde, als naturwissenschaftliche Anwenderin von Erkenntnissen der Vergangenheit, annehmen, dass jedes Mitglied eines bestimmten Fachgebiets die Vereinbarungen innerhalb dieses Fachgebiets teilt und insofern die Schlussfolgerungen, nicht aber die Prämissen ihrer Arbeit betonen. Das hat beim Verlassen des jeweiligen Fachkreises allerdings ganz konkrete, teils negative Auswirkungen auf Lehr-/Lernprozesse. Caspari, Weinrich et al. (2018) und Caspari, Kranz et al. (2018) konnten zeigen, dass in universitären Lehrveranstaltungen der organischen Chemie, der Unterscheidung von Oberflächen- und Tiefenstrukturen eine besondere Rolle zukommt: Verweisen Hochschullehrer bei der Darstellung von Reaktionsmechanismen nur auf Merkmale, die auf der Tafel sichtbar sind, bekommen die Studierenden in Prüfungssituationen Schwierigkeiten, wenn sie diese auf eine verwandte, aber anders situierte Reaktion transferieren müssen. Ein Beispiel aus der Chemie ist die Regel von Markownikow. Sie lautet nicht „Das zu addierende H-Atom bindet an das C-Atom, das schon die meisten H-Atome hat.“, sondern „Das zu addierende Heteroatom bindet an das C-Atom, das das stabilere Carbenium-Ion bildet.“ Diese Verständnishürde stellt Lehrende sowohl an Universitäten, als auch an Schulen vor große Herausforderungen. Auf erkenntnistheoretische Grundlagen gewendet, lösen sich diese Herausforderungen eben nur dann auf, wenn der Existenz einer oder mehrerer Modellkomponenten keine objektiven Wahrheiten zukommen, sondern allein die empirische Angemessenheit das Akzeptanzkriterium (s. o.) darstellt und der hypothetische Charakter von Modellen expliziert wird. Für eine Konkretisierung bedarf es aber weiterer theoretischer Ausarbeitungen, die im Folgenden dargestellt werden.

Allgemeine Modelltheorie nach Stachowiak (1973)

In einer „[...] kritischen Würdigung [...]“ (Stachowiak, 1973, S. 47 ff.) diskutiert Stachowiak die Zurückweisung des Verifikationismus durch Karl Popper (1935). Insofern Stachowiak sich von der streng falsifikationistischen Theorie Poppers abgrenzt und damit ex negativo definiert, ist die *Allgemeine Modelltheorie* nicht ohne dessen Fundament zu bearbeiten. Im Überblick kann Stachowiaks Beschreibung von Modellen durch drei Merkmale verstanden werden (Stachowiak, 1973, S. 131 ff.).

2. Theoretischer Hintergrund

1. Das Abbildungsmerkmal

Modelle seien immer Modelle *von etwas* (vgl. auch Mahr (2008), s.u.). Den Begriff der Abbildung beschränkt Stachowiak nicht auf eine optische Bildhaftigkeit, sondern verallgemeinert ihn im Sinne einer mengentheoretischen Zuordnung von Attributen des Abgebildeten (Original) auf die Abbildung (Modell) (vgl. Stachowiak, 1973, S. 132). Wenn die Farbe eines Modellautos der Farbe des Originals entspricht, dann ist das ein triviales, optisches Beispiel für eine solche Attributzuordnung. Das wird bspw. dann problematisch, wenn Attribute des Originals gar nicht unmittelbar festgestellt werden können: Atome sind per Definition nicht farbig und können auch sonst nicht mit den Sinnesorganen als einzelne Objekte wahrgenommen werden. Das begründet die Stärke der Darstellung durch Stachowiak. Es ist möglich die Originalattribute konstruktiv anzunehmen und auf empirische Kohärenz zu prüfen (s. o.: *konstruktiver Empirismus*).

2. Das Verkürzungsmerkmal

In einem Modell finden sich nicht alle Attribute des Originals wieder. Weil Modelle durch ihre jeweiligen Konstrukteure oder Nutzer einem bestimmten Zweck dienen, sind auch nur die zweckbezogenen Attribute relevant. Ein Automodell im Maßstab 1:1, das für aerodynamische Messungen verwendet wird, benötigt keinen Motor. Ihm fehlt ein wesentliches Attribut (bzw. eine Menge von Attributen), das das Auto-Sein des Originals wesentlich mitbestimmt. Analog dazu dient die Zeichnung eines Salzkristallgitters nicht dem Zweck, den salzigen Geschmack eines Salzkristalls zu modellieren.

3. Das pragmatische Merkmal

Wie oben angedeutet, würden Modelle eine Ersetzungsfunktion ausüben. Sie seien Modelle *von etwas* (vgl. ebenfalls Mahr (2008), s. u.). Außerdem dienen Sie einem bestimmten Zweck im doppelten Sinne (Modelle *für etwas* und *für jemanden*) (vgl. Stachowiak, 1973, S. 133). Das nennt er die dreifache pragmatische Relativierung, die zu Modellen als „[...] Realisation eines (mindestens) fünfstelligen Prädikats [...]“ (Stachowiak, 1983, S. 118) formalisiert und verallgemeinert dargestellt werden kann: „X ist Modell des Originals Y für den Verwender v in der Zeitspanne t bezüglich der Intention Z.“ (ebd.)

Das Modell des Modellseins nach Mahr (2008)

Der Ansatz von Stachowiaks *Allgemeiner Modelltheorie* ist nach Mahr (2008, S. 192) zu stark idealisiert. Prädikatenlogische Attributionen könnten in systemtheoretischen Zusammenhängen nützlich sein, sie würden aber in konzeptioneller Hinsicht zu kurz greifen (ebd.). Außerdem setzt er die Kenntnis *aller* Attribute des Originals und des Modells für eine eindeutige Attributklassenbildung und damit einen zweckdienlichen Original-Modell-Vergleich voraus (vgl. Stachowiak, 1973, S. 132). Damit ist eine Einschränkung formuliert, die, erstens, vor allem auf nicht-technische Zusammenhänge des Modellbegriffs abzielt (jemand steht einer Künstlerin Modell) und die, zweitens, die Beurteilung

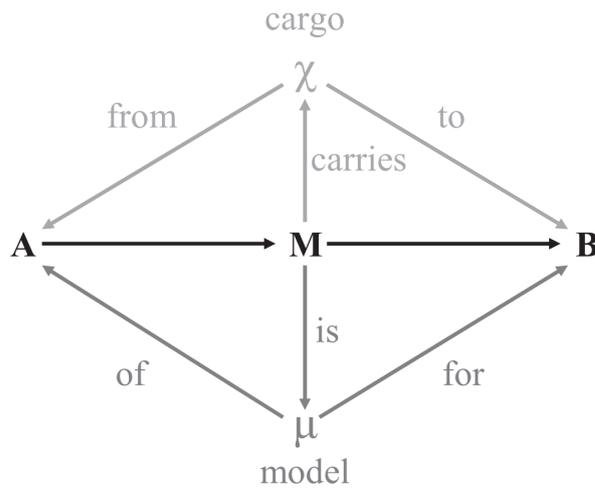


Abbildung 2.1.: Modell des Modellseins nach Mahr (2015).

eines Modells durch Subjekte in den Vordergrund stellt. Wenn ein Modell als fünfstelliges Prädikat (s. o.) zur Bearbeitung einer naturwissenschaftlichen Fragestellung dient, dann kann ein anderes fünfstelliges Prädikat dies passender oder unpassender tun. Beide Modelle können ihre Definition nach Stachowiak vollumfänglich erfüllen, gleichzeitig ist aber ihr Akzeptanzkriterium (empirische Kohärenz, vgl. *konstruktiver Empirismus*) zum modellierten Sachverhalt unterbestimmt. Einen verallgemeinerbaren Ansatz dazu diskutieren Gouvea und Passmore (2017) sowie Passmore et al. (2014) für empirisch-pädagogische Zusammenhänge. Begründet durch diese Herausforderungen, solle der Fokus auf der *Zweck-* oder *Funktionsbestimmung* eines Modells und weniger auf der exakten Zuordnung von Attributen liegen (vgl. Mahr, 2008, 2011, 2015). Das Urteil über die „[...] *Identität als Modell* [...]“ (Mahr, 2008, S. 198) wird einem (auch nicht-physischen) Gegenstand durch ein Subjekt *zugewiesen*. Die Begründung: Allen Modellen liege immer ein herangetragenenes Funktionieren zugrunde (ebd., S. 199) und dies könne nicht durch – dem Gegenstand immanente – Eigenschaften objektiv gefasst werden. Dieses Urteil zum Modellsein stellt Mahr durch die strenge Trennung von modelliertem Objekt und dem Modellobjekt dar, die aus zwei Perspektiven interpretiert werden muss. Diese sind die *Herstellungs-* und die *Anwendungsperspektive*, die als *Modell von etwas* und *Modell für etwas* bei Stachowiak schon angelegt, aber ausschließlich über die prädikatenlogischen Attributionen verhandelt werden. Ein Objekt ist demnach dann als Modell wirksam, wenn es einen zweckgebundenen Übergang zwischen Herstellung und Anwendung gewährleistet. Dazu überträgt das Modellobjekt den *Cargo*, d. h. den Sinnzusammenhang. Abbildung 2.1 illustriert das Urteil über das Modellsein. Herstellung und Anwendung konstituieren sich gegenseitig und sind in alltäglichen wissenschaftlichen oder Lehr-/Lernkontexten zunächst nur analytisch trennbar. Eine reine Anwendungsperspektive ist schwer einzunehmen, weil die Anwendung eines Modells unmittelbar von

2. Theoretischer Hintergrund

ihrem Cargo abhängig ist, der aus der Herstellungsperspektive heraus festgelegt wird. Die Konkretisierung am Bohr'schen Atommodell soll die abstrakten Begriffszusammenhänge verdeutlichen. Gegeben sei die freie Formulierung von Bohrs Postulaten: *a)* Feste Kreisbahnen mit strahlungsfreier Elektronenbewegung, *b)* Quantensprünge nach $f = \Delta E h^{-1}$, *c)* Korrespondenzprinzip mit dem Bahndrehimpuls eines Elektrons als ganzzahliges Vielfaches des reduzierten Plank'schen Wirkungsquantums, sowie zusätzlich das Phänomen der Linienspektren des Wasserstoffatoms. Unterstellt man dem Phänomen, dass ihm das Verhalten submikroskopischer Teilchen zugrunde liegt, muss zwischen beidem ein systematischer Zusammenhang hergestellt werden. Das Modellobjekt eines Atomkerns mit festen Kreisbahnen für die Elektronen und deren Quantensprünge nach Energieeintrag und -abstrahlung fixiert die Postulate (Herstellung, bzw. *Modell von etwas*) und ermöglicht einen hypothetischen oder experimentellen Zugang für Erklärungen und Vorhersagen (Anwendung, bzw. *Modell für etwas*). Gleichzeitig sind es die fixierten und in eine Repräsentation überführbaren Postulate, die als *Cargo* fungieren und damit das Urteil über das Modellobjekt als Modell erlauben. Eine Zeichnung eines Punkts mit umliegenden Kreisen, auf denen sich kleinere Punkte befinden, ist in dieser Hinsicht noch nicht als naturwissenschaftliches Modell qualifizierbar. Ohne den expliziten Bezug zum *Cargo* wird nicht klar, dass eine Absicht zur Modellierung oder Modellanwendung vorliegt.

Mahrs Theorie wird insofern als ein Ansatz von Modellen als Repräsentationen interpretiert. Besonders in der Herstellungsperspektive geht es in seiner Arbeit um den Transfer von Merkmalen des zu modellierenden Sachverhalts auf das Modellobjekt. Damit unterstellt er eine Strukturanalogie als basales Kriterium für das Urteil über das Modellsein. Obwohl dieses subjektive Urteil in der Wissenschaft einen wichtigen Platz hat, kann damit eine naturwissenschaftsdidaktische Darstellung von Modellen noch nicht gewährleistet werden (vgl. Klüver, 2015). Trotz seiner detaillierten Analyse der Zusammenhänge von Herstellung und Anwendung, sowie der Betonung eines pragmatischen Ansatzes, fehlt die Explikation eines solchen in den Ausführungen. Das konstruktivistische Moment ist zwar angelegt, erst die Wendung durch Knuuttila (2005, 2017) und Knuuttila und Boon (2011) wird aber der vorliegenden Arbeit eine solche Explikation hinzufügen: Indem Modelle als epistemische Werkzeuge verstanden werden, können die analytischen Ansätze der allgemeinen Modelltheorie und des Modellseins passend auf Lehr-/Lernprozesse transferiert werden.

Modelle als epistemische Werkzeuge nach Knuuttila (2011)

Alle oben aufgeführten Theorien können als repräsentative Modelltheorien aufgefasst werden (vgl. auch Giere, 2004, 2010). Das bedeutet, sie definieren Modellmerkmale wesentlich über die akkurate Verknüpfung mit den Merkmalen realer Entitäten. Die Struktur und Funktionsweise eines Ellbogens kann über das Basteln eines Modells erschlossen werden, indem die funktionalen Einheiten des Originals in funktionale Einheiten des Modells überführt werden (vgl. Penner et al., 1997). Auf diese Weise könne ein qualitativer Unterschied über die Passung zwischen Modell und Welt hergestellt werden. Knuuttila (2011) problematisiert diesen Ansatz, weil der „Grad der Passung“ (ebd., S. 262,

Übersetzung durch Autor) eben kein Kriterium für ein angemessenes Modell darstellen kann:

„Models contain idealizations, simplifications, approximations, fictional entities and so on, which seem to make them hopelessly inaccurate and distorted representations of the world.“ (Knuuttila, 2011, S. 262 f.)

Diese Dimension der unmittelbaren Strukturverknüpfung (bei Stachowiak (1973): Attributionen; bei Mahr (2008): Modellcargo) von Modell und Original ist nicht nur bei offensichtlichen Beispielen ein Problem. Während eine Übertragung von Strukturmerkmalen eines Atoms auf ein Atommodell prinzipiell nicht herstellbar ist, kann Knuuttilas Kritik auch auf das makroskopische Ellbogenbeispiel übertragen werden. Wenn das Ziel der Herstellung und Nutzung von Modellen der Erkenntnisgewinnung über die Welt dient, dann seien es gerade die Idealisierungen, Vereinfachungen, usw., die zur systematischen Untersuchung eines Sachverhalts beitragen würden (Knuuttila, 2011, S. 267). In diesem Fall wäre die Identifizierung von Modellen als Repräsentationen nicht nur unzutreffend, sondern sogar kontraproduktiv. Die entsprechenden Grenzen verschiedener Modelle tragen nämlich zur multiperspektivischen Untersuchung von Sachverhalten bei und das stellt ein Wesensmerkmal in der gängigen Praxis naturwissenschaftlicher Forschung dar (ebd.). Bohrs erstes Postulat (s. o.) steht beispielsweise im Widerspruch zur Elektrodynamik: Ein geladenes Teilchen bewegt sich nicht strahlungsfrei auf Kreisbahnen. Das ist – löst man sich von der Idee von Modellen als akkuraten Repräsentationen – nach Knuuttila primär nicht erheblich für den wissenschaftlichen Wert des Modells und seines Potentials für Erkenntnisgewinn. Die Begründung: Modelle als zweckorientierte Artefakte enthalten vor allem einen erkenntnistheoretischen Wert, wenn sie einen Möglichkeitsraum eröffnen. Dazu ist die Formulierung von idealisierten Grenzfällen notwendig, die eine Reihe erklärungs- oder prädiktionsmächtiger Modelle auszeichnet (vgl. den Carnot-Prozess bei Knuuttila & Boon, 2011). Eine stärker experimentbezogene Untersuchung dieses Sachverhalts diskutiert Portides (2007). Das bedeutet allerdings nicht zwingend, dass der Repräsentationscharakter überhaupt keinen Wert für ein Modell hat. Mit der systematischen Erkundung von Sachverhalten mittels Modellen ergeben sich mindestens bei der Explikation strukturelle Angleichungen. In diesem Moment ist das Repräsentationsmedium (d. h. der materiale Ausdruck: verbale Beschreibung, Skizze, digitale Animation, etc.) enorm wichtig. Es definiert, wie ein Modellnutzer das Modell manipulieren (s. u.: *konkrete Manipulierbarkeit*) kann. Das Repräsentationsmedium ist getrennt darzustellen vom Repräsentationsmodus (bspw. natürliche Sprache oder Bilder), der auf einer abstrakteren Ebene angelegt ist. Ein beschriebener Sachverhalt, dessen Modus die natürliche Sprache ist, kann in verschiedenen Medien (bspw. Vortrag oder geschriebener Artikel) dargestellt werden (vgl. Knuuttila, 2011, S. 269).⁴

Auch ein sehr abstraktes, mathematisch-physikalisches Modell substituiert reale Gegenstände für Zeichen und ordnen ihnen Relationen zu, die wiederum Relationen zwischen den Originalen entsprechen sollen. Es geht beim Verständnis von Modellen als

⁴Für einen historischen Überblick im Bereich Chemie vgl. Meinel (2008), sowie Klein (2001, 2003). Für die tiefergehende Analyse von Repräsentationsmitteln, jeweils bestehend aus Repräsentationsmodus sowie Repräsentationsmedium, verweist Knuuttila auf Kress und van Leeuwen (2001).

2. Theoretischer Hintergrund

epistemische Werkzeuge folglich nicht darum, ihren repräsentativen Charakter vollständig zurückzuweisen, sondern ihn lediglich nicht als privilegiertes Alleinstellungsmerkmal zu verstehen (vgl. Knuuttila, 2011, S. 263). Außerdem kann in dieser Lesart jedem Modell durchaus ein Repräsentationszweck unterstellt werden. Die Art der Darstellung ist eng verknüpft mit der Möglichkeit des Verständnisses oder der Nutzung des jeweiligen Modells (s. u.: *epräsentative Intransparenz*).

Neben der Abgrenzung zu den vorher genannten Modellkonzepten, formuliert Knuuttila (2011, S. 267 ff.) fünf Charakteristika, mit denen sowohl die Konstruktion, als auch die Nutzung von Modellen beschrieben werden kann. Diese liegen der Inhaltsbeschreibung und Aufgabenoperationalisierung der vorliegenden Arbeit zugrunde und werden auf einen Theorierahmen transferiert, der kontextunabhängige Modellbeschreibungen und -anwendungen erlaubt (vgl. Abschnitt 2.2.2).

1. Beschränktes Design („constrained design“)

Bei der Bearbeitung naturwissenschaftlicher Probleme mithilfe von Modellen ist die Reduktion eines Phänomens ein evidenter Bestandteil (vgl. auch Stachowiak, 1973). Die Annahme, Modelle seien unzureichend, weil sie durch Idealisierung, Vereinfachung und Näherungen gekennzeichnet seien (vgl. Knuuttila, 2011, S. 267), verkennt die prädiktive Stärke von sehr abstrakten, bzw. idealisierten Modellen. Das beschränkte Design bedürfe der Explikation der Vorannahmen. Dabei könne nie notwendigerweise das Ergebnis einer naturwissenschaftlichen Untersuchung auf Basis eines Modells, unmittelbar auf seine Vorannahmen zurückgeführt werden.⁵ Für den Umgang mit dieser Herausforderung nennt Knuuttila (2011) die Triangulation eines Phänomens mit unterschiedlichen Modellen. Das gehört zur alltäglichen Praxis von Naturwissenschaftlern. Klimamodelle, die die Erhöhung der globalen Durchschnittstemperatur durch die anthropogene Erhöhung der CO₂-Konzentration vorhersagen, stehen nicht für sich allein. Sie sind neben anderen theoretisch plausiblen Modellen (nicht-anthropogene Konzentrationserhöhung, kosmische Strahlung, natürliche Schwankung durch atmosphärische Strömungseffekte, etc.) schlicht die beste bekannte Erklärung für alle vorliegenden Beobachtungsdaten. Insofern sind solche Beschränkungen keine Schwäche, sondern gerade die Basis für Erkenntnisgewinnung mit Modellen.

2. Repräsentative Intransparenz („representational non-transparency“)

Ein problematischer Aspekt bei der Interpretation von Modellen ist, dass in Lehr-/Lernumgebungen häufig implizit angenommen wird, die repräsentative Ebene hätte keinen Einfluss auf den Umgang mit einem Phänomen (vgl. Arnold, 2018; Caspari, Weinrich et al., 2018). Diese Annahme ist nicht haltbar. Schon die Änderung der Größenskalierung hat unmittelbare Auswirkungen auf die Möglichkeiten und Beschränkungen bei der Nutzung. Die Darstellung der Reihenfolge der Elemente

⁵Bei einer Blackbox-Untersuchung (vgl. Lederman & Abd-El-Khalick, 2002) sind mehrere gleichwertige Modelle möglich, deren Annahmen unterschiedlich sein können. Dieses Problem findet sich auch bei statistischen Modellen, wie sie bspw. in der vorliegenden Arbeit verwendet werden (Abschnitt 4.4).

im Periodensystem ist unabhängig von der maßstabgerechten Skalierung der Umlaufbahnabstände ihrer Atomradien oder anderer periodischer Eigenschaften. Um das Prinzip des Zusammenhangs zwischen der Elektronegativität einzelner Atome in einer Verbindung und dem entsprechenden Bindungstyp nachvollziehen zu können, ist das ebenfalls nicht notwendigerweise wichtig. Um mithilfe gegebener Atome jedoch das konkrete Reaktionsverhalten vorherzusagen, sind die konkreten Zahlenwerte, ihre Relationen und die räumliche Elektronenstruktur dieser Atome aber von großer Bedeutung. Dann ist die qualitative Repräsentation von Eigenschaften im Periodensystem nicht mehr hinreichend um ein System mit Prädiktionskraft zu sein. Anders als bei Stachowiak (1973), der dies im Sinne des Verkürzungsmerkmals diskutiert, liegt die Betonung hier eher parallel zu Mahr (2008). Er sieht im *Cargo* eines Modells auch eine subjektive Relevanzauswahl des Modellnutzers. Insofern ist die Explikation der Vorannahmen bei der Nutzung eines Modells sowohl von theoretischer, als auch von ganz praktischer Bedeutung.

3. Ergebnisorientierung („results-orientedness“)

Wenn der Zweck eines Modells im Mittelpunkt steht, dann ist eine Repräsentation im Sinne einer Strukturanalogie zweitrangig. Modelle können danach charakterisiert werden als

„[...] specific kinds of entities built in the [*light of their results* and] can be attributed to their holistic *systemic* nature, which distinguishes them from many other scientific representations that often fragment and analyze an object or specimen in minute detail.“ (Knuuttila, 2011, S. 268).

Insofern ein Modell ein System erfolgreich beschreibt oder vorhersagt, ist es nicht notwendig, das zu modellierende System akkurat abzubilden. Ein Beispiel für die Kontextunabhängigkeit eines Modells sind die Lotka-Volterra-Gleichungen aus der theoretischen Biologie. Sie beschreiben die Populationsschwankungen in Räuber-Beute-Beziehungen als System von Differentialgleichungen und wurden erfolgreich für Vorhersagen in volkswirtschaftliche Zusammenhänge transferiert. Dabei wird in den Gleichungen der Räuber durch die Lohnquote ersetzt, die Beute durch die Beschäftigungsquote (Goodwin, 1967). Die zyklischen Bewegungen der beiden Größen verhalten sich dann – bei stabilen Umgebungsbedingungen – mathematisch exakt wie Populationen von Schneehasen und Luchsen. Die Möglichkeit des Transfers wird dadurch hergestellt, dass die abstrakte Idee eines Systems von zwei abhängig voneinander oszillierenden Größen unabhängig von seinem makroskopischen Gegenstand ist (vgl. Knuuttila & Loettgers, 2014). Ähnliches gilt für die unerwartet hohe Ionisierungsenergie in den inneren s-Orbitalen der 4. Periode des PSE: Obwohl die Darstellung von Elektronen als Punktmassen, die sich aufgrund der hohen Kernladung so schnell bewegen, dass eine relativistische anstelle einer klassisch mechanischen notwendig wird, nicht akkurat ist, liefert sie einen hinreichenden Rahmen für die Erklärung und Vorhersage des Phänomens. Das Beurteilungskriterium verschiebt sich so von einer korrekten Repräsentation hin zur Brauchbarkeit des Ergebnisses.

2. Theoretischer Hintergrund

4. Konkrete Manipulierbarkeit („concrete manipulability“)

Versteht man ein Modell als Artefakt, mit dem hypothetische Zusammenhänge formuliert werden können, dann bedarf es der Manipulierbarkeit. Die einzelnen Komponenten eines Modells stehen in bestimmten Bezügen zueinander und es sind diese Bezüge, mit denen der Modellnutzer aus dem hypothetischen Moment eine empirische Prüfung generieren kann. Insofern spielt der Repräsentationsmodus von Modellen eine wichtige Rolle. Ohne eine bestimmte Darstellungsform sind die Modellkomponenten nicht expliziert, dadurch fixiert und in der Folge auch nicht manipulierbar. Knuuttila (2011) deutet diese Externalisierung als

„[...] *scaffolding* which both narrows the space of information search by localizing the most important features of the object in a perceptually salient and manipulable form, and enables further inferences by making the previously obscure or scattered information available in a systematic fashion.“ (ebd., S. 269)

5. Plurale Rechtfertigung („distributed justification“)

Es ist in einer oberflächlichen Näherung durchaus möglich, Modelle allein als Repräsentationen, d. h. Strukturanaloga aufzufassen. Auch das Beispiel von der Übertragung der Lotka-Volterra-Gleichungen auf Wirtschaftszusammenhänge kann so interpretiert werden. Die Lohnquote wird dann eben *repräsentiert* von Symbolen in einer Teilgleichung, das zyklische Auf und Ab des Graphen im entsprechenden Diagramm *repräsentiert* dann die durchschnittliche Schwankung einer Geldmenge. Das funktioniert aber für die Nützlichkeitsbeurteilung nur noch bedingt. Knuuttila (2011) fragt, welche Hinweise für die Beurteilung eines Modells vorliegen würden, wenn es nur darstellt was an bereits vorliegendem Wissen hineingesteckt wurde. Wenn Modelle ausschließlich für die Darstellung von Vorhandenem genutzt würden, dann ginge das hypothetische Moment verloren und das ist unplausibel, was am Beispiel eines submikroskopischen Teilchens gezeigt werden kann. Wir können *prinzipiell* Ort und Impuls eines submikroskopischen Teilchens nicht bestimmen. Es bedarf also einer Formulierung des *Es-Wäre-Möglich*, dessen Konsequenzen anschließend auf die Passung mit dem Phänomen geprüft werden. Damit sind historisch gewachsene Konsensmodelle in ihrem Wert überhaupt nicht berührt. Es wird, im Gegenteil, sogar der Wert des stets revidierbar zu haltenen naturwissenschaftlichen Vorgehens gestärkt. Das funktioniert in allen Größenordnungen, die per Messungen zugänglich sind, d. h. von der atomaren bis auf die kosmologische Ebene. Versteht man bspw. die Wechselwirkung von Zeit, Raum und Masse nach der Relativitätstheorie als Modell, das die Bewegung von Körpern vorhersagen kann, dann ist es genau diese hypothetische, vorläufige Vorgehensweise, die relevanten Erkenntnisgewinn liefert und die Teil aktueller Forschung an der Grenze menschlicher Technologie und Erkenntnisfähigkeit ist (vgl. Herrmann et al., 2018). Zu einer pluralen Rechtfertigung eines Modells gehören in der Folge auch die jeweilige Domäne (bspw. Chemie), etablierter Wissenskonsens (auf einer Fachkonferenz zu analytischer Chemie wird niemand die Avogadro-Konstante hinterfragen), For-

malismen und empirische Daten. Ein Modell wird dann in der Schnittmenge dieser und vieler weiterer – auch fachfremder – Perspektiven seinen Wert erhalten.

Zusammenfassung

Für das Nutzen von Modellen wurden verschiedene erkenntnistheoretische Ansätze dargestellt und deren Bezüge zur Erkenntnisgewinnung im Sinne des hypothetisch-deduktiven Denkens formuliert. Ein Schwerpunkt ist die Deutung von Modellen als epistemische Werkzeuge (Knuuttila, 2011; Knuuttila & Loettgers, 2014). Werkzeuge werden von jemandem für einen bestimmten Zweck angefertigt und verwendet. Dadurch wird in diesem Verständnis die subjektive Nutzung von Modellen besonders stark in den Vordergrund gestellt. Auch die Unabhängigkeit des zu untersuchenden, naturwissenschaftlichen Sachverhalts von dem Modell, das zu dieser Untersuchung verwendet wird, erhält durch diese Begriffswahl einen Schwerpunkt. Die Komponenten eines Modells brauchen nicht möglichst genau der Struktur des Sachverhalts zu entsprechen und aus diesem Grund werden Modelle auch nicht mit Repräsentationen gleichgesetzt. Gleichzeitig ist in naturwissenschaftlichen Modellen durchaus ein repräsentatives Moment enthalten, das sich bei Stachowiak (1973) im Abbildungsmerkmal niederschlägt. Mahr (2008, 2015) stellt diesen Bezug über das *Modell des Modellseins* her. Die Zweckbestimmung eines *Modellobjekts* wird durch zwei Perspektiven bestimmt. Dabei ist die Herstellungsperspektive, d. h. das *Modell von etwas*, im Wesentlichen davon bestimmt, Zusammenhänge eines Sachverhalts auf die Darstellung im Modellobjekt zu übertragen. Die Anwendungsperspektive spielt für die vorliegende Arbeit die dominante Rolle. Beim *Modell für etwas* bedarf es des Urteils über die Eignung des Modellobjekts für eine naturwissenschaftliche Untersuchung, hier: hypothetisch-deduktiver Erkenntnisgang (vgl. auch *Modelle als fünfstellige Prädikate* bei Stachowiak, 1973, 1983). Tabelle 2.1 fasst den Abschnitt überblicksweise zusammen. Die Frage, was genau ein Modell in ontologischer Hinsicht ist,

Tabelle 2.1.: Überblick der ausgewählten theoretischen Rahmungen.

Theorie	Epistemologie	Bezug zur Studie
wissenschaftl. Realismus	objektivistisch	orientierungsbildend, Atome als reale Objekte
konstruktiv. Empirismus	subjektivistisch	Phänomene als Primat der Naturwissenschaften
allg. Modelltheorie	systemtheoretisch	formale Rahmgebung für Modellnutzung
Modell des Modellseins	subjektivistisch	Differenzierung der Zweckorientierung
Modelle als Werkzeuge	subjektivistisch	Kritik am Repräsentationsbegriff, sprachl. Aspekte

Anm.: –

2. Theoretischer Hintergrund

tritt zugunsten seiner Erklärungs- oder Prädiktionskraft in den Hintergrund (vgl. auch Kind & Osborne, 2017; Krüger et al., 2018). Den Rahmen für dieses Verständnis von Modellen im Kontext naturwissenschaftsdidaktischer Forschung und Anwendung bildet ein moderat konstruktivistischer Ansatz. Für eine nützliche Beschreibung der Welt ist es unerheblich, ob Atome tatsächlich die exakte Struktur besitzen, die ihnen durch ein Modell zugeschrieben wird. Wenn diese vorbehaltliche Zuschreibung aber widerspruchsfrei zur Konsequenz hat, dass die beste Erklärung für Phänomene die Existenz kleinster Teilchen ist, dann können daran Erkenntnisse gewonnen und der Umgang mit ihnen erlernt werden. Zur vertiefenden Diskussion dieses Standpunkts vergleiche Devitt (2006) und van Fraassen (1980) sowie Giere (2004, 2010).

2.2.2. Modellkomponenten zur chemiespezifischen Modellnutzung

Dieser Abschnitt beschreibt einen Ansatz zur Konkretisierung des oben aufgeführten, erkenntnistheoretischen Rahmens. Wenn die oben beschriebenen, erkenntnistheoretischen Grundlagen auf eine spezielle Aufgabenstellung angepasst werden sollen, dann bedarf es einer begrifflichen Verbindung. Für den Übergang von den epistemologischen Überlegungen hin zur empirischen Übersetzung in problemhaltige Umgebungen, werden vier Modellkomponenten eingesetzt, die die strukturelle Basis jedes in dieser Arbeit verwendeten Modells darstellen. Sie werden als *Elements*, *Relations*, *Operations*, und *Rules* benannt und sind ursprünglich in mathematikdidaktischen Arbeiten formuliert worden (Lesh et al., 2003; Lesh et al., 2000; Lesh & Zawojewski, 2007). In der empirischen chemiedidaktischen Forschung sind sie ebenfalls schon eingesetzt worden (vgl. Bodner & Briggs, 2005; Schwarz et al., 2009), allerdings ohne dass bisher eine systematische Untersuchung zu ihrer lernwirksamen Nutzbarkeit oder ihrer empirischen Unterscheidbarkeit stattgefunden hätte.

Einerseits wird hier die Nutzung der vier Komponenten im Hinblick auf eine starke Subjekt- und Momentgebundenheit untersucht. Es wäre also zu erwarten, dass dieser Prozess sehr offen verläuft und verschiedene Nutzungsstrategien und -muster vorliegen. Andererseits wird eine relativ enge, durch die Bildungsstandards motivierte Perspektive auf die Modellnutzung geworfen, indem Erkenntnisgewinnung als Durchlaufen eines hypothetisch-deduktiven Dreischritts (Abschnitt 2.1.1) erfasst wird. Hinzu kommt eine große Heterogenität des jeweiligen Fachkontexts und der Zielstellungen der Aufgaben (Abschnitt 2.3). Theoretisch sollte der fachliche Kontext in den Hintergrund treten. Denn wenn die Annahme stimmt, dass das Nutzen von Modellen eine abstrakte Fähigkeit ist, dann sollte sie über verschiedene Situationen hinweg anwendbar sein. In diesem Fall ist aber eine Standardisierung beim Umgang mit Modellen notwendig. Während der hypothetisch-deduktive Dreischritt bereits empirisch für Lehr-/Lernprozesse beschrieben wurde, steht die Verknüpfung mit einem verallgemeinerbaren Begriffssystem zur Modellnutzung noch aus. Könnte gezeigt werden, dass die Beschreibung von zu nutzenden Modellen in chemischen Kontexten standardisierbar ist, dann wären die vielen unterschiedlichen, ontologischen Kategorien von Modellen immernoch interessant für erkenntnistheoretische Diskussionen. Für einen Transfer in Lehr-/Lernprozesse wäre aber eine begriffliche Sparsamkeit gewonnen, die die Lehrenden und Lernenden entlastet. Die

vier Modellkomponenten sind in dieser Arbeit für die systematische, kontextunabhängige Modellnutzung vorgesehen und werden folgend mit verschiedenen Einzelbeispielen illustriert. Nach der Darstellung der einzelnen Komponenten wird ein Beispiel gegeben, bei denen alle vier Komponenten konsistent aufeinander bezogen werden.

Kleinste Sinneinheiten („Elements“)

Als *Elements*⁶ eines Modells werden die kleinsten, sinntragenden Einheiten bezeichnet. Sie stellen die begriffliche Grundlage bei der Modellnutzung dar und sind deklarativ. Das bedeutet, dass zur Nutzung eines Modells die subjektive Entscheidung darüber getroffen werden muss, welche *Elements* relevant im Sinne der jeweiligen Aufgabenstellung sind. Wenn die Strukturformel eines organischen Moleküls als Modell verstanden wird, dann könnten beispielsweise die einzelnen Atome, oder auch die bindenden Elektronen als *Elements* benannt werden. Beide könnten, je nach Zielstellung, gleichartig als Kugelteilchen benannt, oder auch bewusst unterschiedlich als Kugeln (Atome) und Stäbe (Elektronenpaare) beschrieben werden.

Relationen („Relations“)

Die *Relationen* setzen die Identifikation der *Elements* voraus, weil sie verschiedene Beziehungen zwischen diesen beschreiben. Sie können räumliche Beziehungen ausdrücken, indem den *Elements* verschiedene Volumina (bzw. Ausdehnungsradien) zuweisen. Auch Winkel zwischen zwei Bindungen sind *Relationen* in diesem Sinne. Darüber hinaus können auch physikalische Größen (sowohl qualitativ als auch quantitativ) zu diesem Zweck genutzt werden. Werden Moleküle als Kugelteilchen dargestellt, so gibt es eine Verteilung der kinetischen Energie, sodass diese Größe als *Relation* dienen kann. Manchen Teilchen ist dann *mehr*, anderen *weniger* Energie zugeordnet.

Operationen („Operations“)

Diese Komponente beschreibt die Veränderung der *Relationen* bei der Nutzung von Modellen. Weil Modelle hypothetische Werkzeuge sind, ist es zunächst unerheblich ob diese Veränderung ein empirisches Relativ besitzt. Unter der Annahme, dass ein Elektron in einer Atomhülle sich auf einem bestimmten Energieniveau befindet, kann der Wechsel auf ein höheres Energieniveau durch Zufuhr thermischer Energie als *Operation* verstanden werden. Die Veränderung von Bindungswinkeln bei einer nukleophilen Substitution zweiter Ordnung (*Walden-Umkehr*) ist ein weiteres Beispiel, bei dem eine solche Veränderung der *Relationen* modellhaft beschrieben werden kann.

Regeln („Rules“)

Die drei genannten Modellkomponenten müssen für eine konsistente Beschreibung, Erklärung oder Vorhersage eines Sachverhalts systematisch benannt und/oder miteinander

⁶Da der Begriff *Element* eine feste Bedeutung in der Chemie hat, wird er im Lauf der vorliegenden Arbeit durch *kleinste Sinneinheiten* ersetzt (vgl. auch Abschnitt 5.1).

2. Theoretischer Hintergrund

der verknüpft werden. Zu diesem Zweck werden die *Regeln* verwendet, die hier mit der Anwendung von Wahrheitstafeln durch Existenzaussagen (Quantoren) und Aussageverknüpfungen (Junktoren) aus der klassischen Logik identifiziert werden.⁷ Der Benennung eines *Elements* beispielsweise, liegt damit formal die Anwendung des Existenzquantors (\exists Atom) zugrunde. Eine *Relation* kann folgendermaßen beschrieben werden: Es existieren zwei voneinander verschiedene Energieniveaus und ein Elektron, das sich immer (nur) auf einem der beiden Energieniveaus befindet. Für die vorliegende Arbeit wird keine tiefergehende Diskussion über Syllogismen oder andere Aspekte der Aussagenlogik geführt. Sie ist ein Mittel zur Systematisierung der Modellnutzung und soll durch ihre basale Struktur der Vereinheitlichung von Begriffen dienen (vgl. auch Abschnitt 2.1.1). Die empirische Nutzbarkeit ist, zumindest für Aussagen der Form *Wenn-Dann*, qualitativ bereits für naturwissenschaftlichen Unterricht beschrieben worden (Ergazaki & Zogza, 2005).

Beispiel

Für die Verbrennung eines organischen Stoffs, werden die Moleküle dieses Stoffs und elementarer Sauerstoff benötigt. Wenn die jeweiligen Atome die kleinsten Sinneinheiten sind, dann können die Bindungskonfigurationen (Abstand, Winkel, Mehrfachbindungen) als deren Relationen aufgefasst werden. Die Bindungsumordnung d. h. die Operation, unterliegt Regeln: Wenn Methan mit Sauerstoff reagiert, dann entstehen Kohlenstoffdioxid und Wasser. Die Komplexität oder Detailtiefe ist dabei je nach Zielsetzung wählbar. Im genannten Beispiel wird nur auf die Bildung eines neuen Stoffs durch Atomumordnungen Bezug genommen. Alternativ könnte auch eine energetische Betrachtung entsprechend modelliert werden. Dann wären die Standardbildungsenthalpien der Moleküle die kleinsten Sinneinheiten, die Differenzen bzw. Summen der Standardbildungsenthalpien wären die Relationen und die jeweiligen Änderungen die Operationen. Wenn eine exotherme Reaktion abläuft, dann ist die Differenz der Standardbildungsenthalpiesummen zwischen den Produkten und Edukten kleiner als Null. In diesem Fall steht hinter der frei formulierten Regel sogar ein kanonisches Fachgesetz, nämlich der Satz von Hess.

Zusammenfassung

Die starke Formalisierung des Beschreibungsrahmens mit Hilfe der Modellkomponenten hat für diese Arbeit drei Vorteile. Erstens funktioniert die klassische Logik für beliebige Aussagen. Da die Modellnutzung als hypothetisch verstanden wird und die epistemologische Grundannahmen hier konstruktivistisch sind, ist der Modellnutzer nicht darauf angewiesen, dass es empirisch strukturanaloge Entsprechungen des Modells gibt. Das ist passend für klassische Logikoperationen, bei denen es nicht darauf ankommt, dass die zu untersuchenden Aussagen wahr sind, sondern nur, dass sie wahr wären, wenn die postulierten Annahmen zutreffen würden (vgl. Abschnitt 2.1.1). Unter der Bedingung,

⁷Gebräuchliche Quantoren sind der *Existenzquantor* (*Es-Gibt-Mindestens*, \exists) und der *Allquantor* (*Für-Alle-X-Gilt*, $\forall x$). Gebräuchliche Junktoren sind das *Nicht* (\neg), das *Und* (\wedge), das *Oder* (\vee) und das *Wenn-Dann* (\rightarrow).

dass Photonen als massebehaftete Teilchen beschrieben werden, kann der photoelektrische Effekt zugänglich erklärt und für technische Anwendungen quantifiziert werden. Es ist dabei zunächst von untergeordneter Bedeutung, dass es bei dieser Beschreibung von Photonen zur Schwierigkeit kommt, dass der Wellencharakter des Lichts damit nicht vereinbar ist. Darauf aufbauend kann, zweitens, theoretisch angenommen werden, dass der jeweilige Themenbereich keinen unmittelbaren Einfluss auf die Bearbeitung eines Problems hat. Drittens können mit dem Beschreibungsrahmen über die *Regeln* Wahr-Falsch-Aussagen generiert werden, die in der empirischen Umsetzung der latenten Fähigkeitsmessung als korrekt/unkorrekt codiert werden (vgl. Abschnitt 4.4). und die die Modellnutzung mit dem hypothetisch-deduktiven Verfahren verknüpfen. Damit ist die theoretische Grundlage formuliert, um relativ komplexe Aufgaben zu konstruieren, die gleichzeitig einer effizienten quantitativ-statistischen Auswertung zugänglich sind.

2.3. Chemiespezifische Themen für die Modellnutzung

In der Chemie ist der Begriff des *Teilchens* der primäre Bezugspunkt und der Umgang mit Begriffen und Zusammenhängen auf der submikroskopischen Ebene scheint ein kritisches Moment der Inhaltsbewältigung in der Chemie zu sein (Sumfleth & Nakoinz, 2019). Alle fünf Aufgabenthemen werden theoretisch darauf zurückgeführt und erhalten damit ihre fachspezifische Bedeutung. Auf Basis der Annahme einer diskontinuierlichen Welt und den dafür vorliegenden Evidenzen, leiten sich aus den entsprechenden Modellen Beschreibungen, Erklärungen und Vorhersagen für Phänomene ab. Tabelle 2.2 zeigt die fünf inhaltlichen Bezüge und deren curriculare Einordnung im Überblick. Aus

Tabelle 2.2.: Fachinhaltliche Themen der Studienaufgaben.

Thema	Subdomäne innerhalb der Chemie	übliche Verortung in Schulcurricula
Metall- & Ionenbindung	chemische Bindungen, Stoffeigenschaften	Anfangsunterricht, 7.-8. Klasse
Destillation	Stofftrennverfahren, Industrieprozesse	10. Klasse, wenn bezogen auf Rohöl/OC
Wasser als Dipol	Löslichkeit, polare Atombindung	vertiefende Aspekte chem. Bindungen, 9.-11. Klasse
Reaktionsgleichungen	universell	alle Klassenstufen
Strukturformeln	Reaktionsmechanismen, Eigenschaft folgt Struktur	10. Klasse bis gym. Oberstufe

Anm.: –

der Kenntnis über den Unterschied zwischen der Teilchenstruktur eines Metalls und der

2. Theoretischer Hintergrund

Teilchenstruktur eines Salzkristalls, ergeben sich jeweils andere Konsequenzen für die Eigenschaften der entsprechenden Stoffe. Weil ohne Materialwissenschaften, die sich auf chemischen Strukturen beziehen – z. B. in der Bauteilherstellung – eine moderne Gesellschaft nicht denkbar ist, wird dieses Thema als relevant angesehen. Die Erforschung und Herstellung von neuen Materialien mit speziellen Eigenschaften ist integraler Bestandteil der modernen Naturwissenschaften. Viele aktuelle Herausforderungen in Hochtechnologiesegmenten wie der Raumfahrt oder dem Energiesektor sind empfindlich von der Entwicklung bspw. extrem hitzeresistenter oder leichter und gleichzeitig stabiler Materialien abhängig. Für einen Teil der Aufgaben dieser Studie wird der Vergleich zwischen einer Metall- und einer Ionenbindung vorgenommen, deren Unterschiede chemietypisch auf die Elektronenstruktur der Gitterbestandteile zurückgeführt werden.

Großtechnische Prozesse übertragen physikalisch-chemische Prinzipien auf die Produktion bedeutsamer Roh- und Fertigprodukte. Zu diesen Prozessen gehört die Destillation von Rohöl, die die Kunststoff-, Pharmazeutika- und Treibstoffherstellung überhaupt erst möglich macht. Es werden milliardenschwere Umsätze mit der Gewinnung und Verarbeitung von Rohöl erzielt und die aktuellen Debatten um die zukünftigen Klimaveränderungen sind untrennbar mit der Verarbeitung und Nutzung fossiler Energieträger verknüpft. Außerdem ist die Öltraffination, d. h. die Gewinnung von organischen Basischemikalien im Kilotonnenmaßstab weit über die Frage zukünftiger Antriebstechnik und Energiegewinnung hinaus relevant.

Wenn die Chemie sich mit den Wechselwirkungen kleinster Teilchen beschäftigt, dann sind Untersuchungen zum Löslichkeitsverhalten von Stoffen ein wichtiger Teil dieser Wissenschaft. Egal ob es um den Wohnungssputz oder um die Beurteilung der Umweltverträglichkeit von Chemikalien geht: Die Arbeit mit der Löslichkeit erlaubt die Einschätzung von Wechselwirkungen verschiedener Stoffe. Darum handelt es sich um ein bedeutsames Thema, das über den rein inhaltlichen Bezug von Wasserlöslichkeit hinaus geht.

Reaktionsgleichungen sind archetypische Repräsentanten der Chemie. Als übergeordnetes Formalsystem erlauben sie, bei bekannten Stoffmengenverhältnissen der entsprechenden Reaktionen, vor allem die Quantifizierung von Stoffumsätzen. Sie verknüpfen die phänomenologische Ebene der Massenbestimmung mit der modellhaften Ebene der einzelnen Teilchen und – sofern man weiß wie mit ihnen umzugehen ist – erleichtern die Untersuchung und Darstellung einer chemischen Reaktion.⁸ Aus diesem Grund ist es unumgänglich, sich mit ihnen zu beschäftigen.

Während Reaktionsgleichungen mit Summenformeln die Quantifizierung von Stoffumsätzen erlauben, sind Strukturformeln nützlich, um konkrete Reaktionsverläufe und Wechselwirkungen zwischen Teilchen zu beschreiben. Die räumlich-geometrische Darstellung der Struktur eines Teilchens und seiner Reaktionspartner führt zu Annahmen über sterische Hinderungen oder bspw. dem Induced-Fit-Prinzip aus der Biochemie. Umgekehrt können aus der Strukturbestimmung von Syntheseprodukten Rückschlüsse über den Verlauf einer Reaktion gemacht werden: Wenn ein Molekül enantiomerenrein

⁸Es wäre extrem umständlich, würde in einer fachlichen, möglicherweise sogar schriftlichen Darstellung ständig darauf verwiesen, dass ca. $12 \cdot 10^{23}$ Teilchen Natrium mit ca. $6 \cdot 10^{23}$ Chlor-Doppelmolekülen miteinander zu ca. $12 \cdot 10^{23}$ Natriumchlorid reagieren.

aus einer Reaktion hervorgeht, dann sind bestimmte Reaktionswege und -geometrien ausgeschlossen.

2.3.1. Historische Einordnungen

Die Vorstellung einer diskontinuierlichen Welt, die aus kleinsten, unterscheidbaren Teilchen besteht, geht nachweislich bis in die Antike zurück (Weyer, 2018, S. 44 ff.). Leukipp und Demokrit wirkten als sogenannte Vorsokratiker in einer Zeit, als Experimente⁹ für Naturforscher noch nicht selbstverständlich zum Handwerk gehörten. Allein durch Denken und Argumentieren, so die Vorstellung damals, konnten Menschen alle erforderlichen Erkenntnisse gewinnen. Darum blieben die drei Grundannahmen von Leukipp und Demokrit historisch lange strittig: Erstens sei die Welt aus Teilchen aufgebaut. Bis dahin ergab sich noch gar kein Widerspruch zu bestehenden Vorstellungen über die Natur der Materie. Zweitens, daran entzündete sich der Streit, gäbe es eine Untergrenze der Teilbarkeit. Das letzte, nicht weiter teilbare Teilchen wurde als das bis zum heutigen Tage bekannte Atom bezeichnet (*atomós*, gr. das Unteilbare). Es gibt noch einen dritten Punkt, dessen Diskussion bis heute in der Fachdidaktik unter dem Begriff *horror vacui* verhandelt wird und der besonders für jüngere SchülerInnen eine große, gedankliche Herausforderung darstellt (vgl. Abschnitt 2.2.1). Es ist nicht trivial anzunehmen, dass es einen leeren Raum gibt, in dem sich die volumenbehafteten Atome befinden, zwischen denen keine weitere Materie existiert. Der Streit zwischen den Anhängern einer Kontinuumsannahme und den Vertretern einer diskontinuierlichen Welt¹⁰ wurde bemerkenswerterweise erst in der Moderne endgültig beigelegt. Die Atomtheorie nach Dalton konnte erst zum allgemein anerkannten Konsens werden, als die experimentelle Methode zum prozeduralen Wesenskern der Naturwissenschaften wurde. Dem Konsens über die Atomtheorie folgt historisch die Frage nach den inneren Strukturen eines Atoms, die ebenfalls Auswirkungen auf die makroskopischen Eigenschaften eines Stoffs haben.

Bindungsarten als Grundlage für Stoffeigenschaften

Lange bevor ihre zugrunde liegende Elektronenstruktur bekannt wurde, wurden Stoffeigenschaften und deren Veränderungen durch chemische Reaktionen systematisch genutzt. Prähistorisch ist die nicht zufällige Härtung von Holz durch Feuer nachweisbar, auch wenn eine teilchenbasierte Theoriebildung zu dieser Zeit unwahrscheinlich erscheint.

⁹Experimente werden hier, im Kontrast zu anderen Arbeitsweisen in den Naturwissenschaften, streng definiert als ein System von Arbeitstechniken, deren Kern die Anwendung von *Variablenkontrollstrategien* (VKS) darstellt. Vergleiche dazu Schwichow und Nehring (2018).

¹⁰Exakterweise muss hier darauf hingewiesen werden, dass die Annahme einer aus Teilchen bestehenden Welt und einer diskontinuierlichen Welt nicht identisch sind. Die Modellierung der kleinsten Materiebausteine als Wellen ist seit dem Aufkommen der Quantenmechanik das abstrakteste, aber erklärungsmächtigste Werkzeug, sowohl für die Grundlagenforschung als auch für technische Anwendungen. Diese Materiewellen sind aber gequantelt, d. h. sie nehmen definierte Zustände mit dem Planck'schen Wirkungsquantum als Größenordnung für die Mindestdifferenz dieser Zustände ein. Dazwischen existiert nach wie vor nichts Messbares.

2. Theoretischer Hintergrund

Das Kupferbeil des tragisch-berühmten *Ötzi* belegt eine systematische Metallbearbeitung und die Verhüttung, d. h. die Anwendung redoxchemischer Verfahren, von Kupfererzen reicht bis deutlich vor das Jahr 3000 v. Chr. in Mesopotamien zurück (Weyer, 2018). Allein die historischen Zuschreibungen einer Bronze- und einer Eisenzeit sind eindrückliche Nachweise für die Bedeutung der antiken Materialwissenschaften.

Destillation als Beispiel für Stofftrennverfahren

Das Verfahren der Destillation, d. h. der Trennung von Stoffen nach ihrem Siedepunkt, ist schon für die Antike dokumentiert (Weyer, 2018, S. 109 ff.). Im Gegensatz zum Umgang mit Teilchenmodellen, ist der Kontext der Destillation zunächst makroskopisch-technischer Natur. Es geht, in der modernen Terminologie, um die Reinigung und Analyse von Stoffen und nicht um die Frage der grundsätzlichen Zusammensetzung der Welt, obwohl in historischer Hinsicht beides eng miteinander zusammenhängt. Die Suche nach der *quintia essentia*, d. h. dem alles zusammenhaltenden Urstoff, war bei den Philosophen der Antike und den Alchemisten des Mittelalters ein zentrales Anliegen zum Erkennen der Welt (Suhr, 2017). Dazu gehörte auch die Isolierung von Ethanol aus Wein oder anderen Gärprodukten, um daraus die Grundlage für Heilmittel herzustellen. Es ging also darum, einen direkt nutzbaren Stoff zu erhalten. Ebenso war das sogenannte *Griechische Feuer* als antike Kriegswaffe ein elaboriert hergestelltes und gleichzeitig rein zweckorientiertes Produkt auf Erdölbasis (Weyer, 2018, S. 88 f.).

Wasser als Vertreter von Dipolmolekülen

Die Frage danach, ob und wie bestimmte Stoffe systematisch ineinander löslich sind, wurde historisch erstmals vor allem für die Extraktion von Duftstoffen oder Medikamenten relevant, d. h. für die Bereitstellung von teuren Luxusgütern. Die komplexe, weil teils mehrstufige, Extraktion von ätherischen Ölen aus Pflanzen ist bis mindestens 2000 Jahre v. Chr. nachweisbar (vgl. Weyer, 2018, S. 25 f.). Insofern ist hier auch ein Zusammenhang mit der Destillation (s. o.) gegeben: Spätestens im Mittelalter gewann man Ethanol aus Gärmischungen, um es anschließend für die Extraktion von schlecht wasserlöslichen Stoffen verwenden zu können. Gerade weil es so selbstverständlich erscheint, dass Wasser mit seinen besonderen Eigenschaften die Grundlage für Leben darstellt, kann leicht übersehen werden, dass seine Erforschung an der Grenze des wissenschaftlich und technisch Machbaren aktueller nicht sein könnte. Unter der Annahme, dass es erstrebenswert wäre auf anderen Planeten als der Erde Menschen anzusiedeln oder nach erdfremdem Leben zu suchen, ist das Vorhandensein von Wasser eines der wichtigsten Klassifikationsmerkmale für Himmelskörper. All das basiert auf einem grundlegenden Verständnis seiner chemischen Eigenschaften, die auf seine submikroskopische Struktur zurückgehen.

Reaktionsgleichungen als Umordnung von Teilchen

Mit der Einführung einer Schreibweise, die heute systematisch mit dem Begriff der *Summenformeln* zusammengefasst und bspw. von Strukturformeln (s. u.) abgegrenzt wird,

trug der schwedische Chemiker Berzelius einen wichtigen Baustein in der Formalisierung der chemischen Fachsprache bei. Klein (2001) arbeitet die Nutzung von Summenformeln als Modelle für die Aufklärung der Reaktion von Ethanol mit Schwefelsäure zu Diethylether auf. Beschrieben sind die Argumentation an Teilchenumordnung per Summenformeln in Reaktionsgleichungen und das Ausschlussprinzip für mögliche Zwischen- und Endprodukte anhand der entsprechenden Massen. In dieser Hinsicht werden Summenformeln als „[...] paper tools [...]“ (ebd., S. 7) bezeichnet. Die Manipulation dieser Chemie auf dem Papier

„[...] did not follow any logical or methodological rule, but were a kind of tinkering, constrained only by the algebraic syntax of formulas and by long-term concepts, like the concepts of displacement reaction and binary constitution.“ (ebd., S. 13)

Sie sind danach seit dem frühen 19. Jahrhundert hypothetische Werkzeuge, die im Zusammenspiel bspw. mit dem Satz von der Erhaltung der Masse, empirisch prüfbare Vorhersagen ermöglichen.

Strukturformeln als Darstellungsform einzelner Moleküle

Historisch schließt die Nutzung von Strukturformeln an die Etablierung der Summenformeln an. Mit der Erweiterung der Chemie, von einer rein analytischen zu einer auch synthetischen Wissenschaft in den 1840er Jahren in London, erhielt das Verständnis der räumlichen Ausrichtung organischer Moleküle einen enormen Bedeutungsschub (Meinel, 2008):

„Zum Friday Evening Discourse in der London Royal Institution [...] überraschte [August Wilhelm] Hofmann sein Publikum am 7. April 1865 mit einem [...] Modell, dessen Elemente einem beliebten Gesellschaftspiel entstammten: Krocket-Bälle waren Hofmanns Atome[.]“ (ebd., S.224 f.)

Der kompetente Umgang mit Strukturformeln für die Bewertung von Sachverhalten geht allerdings weit über die reine Fachwissenschaft oder ferne, historische Kuriositäten hinaus. Roth (2005) arbeitet den Contergan-Skandal in Deutschland aus einer fachlichen Perspektive auf und demonstriert Missverständnisse und Grenzen bei der Nutzung von Strukturformeln an der Grenze zwischen Chemie und Gesellschaft.

2.3.2. Fachliche Beschreibungen

Bindungsarten

Der Vergleich von Metall- und Ionenbindung kann mit dem Vergleich des Brechens eines Salzkristalls und der Verformung eines Metallkristalls bearbeitet werden. Dazu sind die jeweilige Gitterstruktur und die Spezifika der Bindungen notwendig. Die Bindung im NaCl lässt sich mit einem vollständigen Übergang des Außenelektrons eines elementaren Natriumatoms auf die freie Außenschale eines elementaren Chloratoms formalisieren. Es

2. Theoretischer Hintergrund

liegt demnach eine Redoxreaktion vor. Der vollständige Übergang des Elektrons als Ladungsträger führt zur Ausbildung eines Natrium- und eines Chloridions, die sich anschließend elektrostatisch anziehen. Die Stärke der Anziehung wird über das Coulomb'sche Gesetz quantifiziert (Janiak & Riedel, 2015). Die Anziehungskraft ist demnach proportional dem Produkt der Ionenladungen und umgekehrt proportional dem Quadrat des Abstandes zwischen den Ionen. Sie wirkt gleichmäßig in alle Raumrichtungen und sorgt so für die Ausprägung eines Ionenkristalls, bei dem im Fall von Natriumchlorid jedes Anion von sechs Kationen und vice versa umgeben ist.

Bei einer Metallbindung liegt eine andere Struktur vor. Sie kann beschrieben werden, indem die Metallatome mit allen Elektronenschalen außer der äußersten als Kugeln modelliert werden. Die positiv geladenen Metallrümpfe sind in einer möglichst dichten Packung angeordnet und die Valenzelektronen können sich frei durch die gesamte Gitterstruktur bewegen (sog. Elektronengasmodell). Die gegenseitige Abstoßung der positiven Metallrümpfe wird kompensiert durch die negative Ladung des Elektronengases. So ist es auch zu erklären, dass Metalle bei einer Krafteinwirkung verformt werden, Salze jedoch zerbrechen: Werden bei einem Metall die Gitterebenen gegeneinander verschoben, hält das frei bewegliche Elektronengas die neue Anordnung der Gitterkugeln immernoch zusammen. Im Fall des Salzes treffen bei einer solchen Verschiebung Ionen gleicher Ladung aufeinander, diese stoßen sich ab, die Struktur wird zerstört und es bilden sich scharfe Bruchkanten an den entstehenden Kristallsplintern.

Destillation

Erhitzt man ein homogenes Flüssigkeitsgemisch, so wird es bei konstanter Wärmezufuhr irgendwann beginnen zu sieden. Auf der Teilchenebene lässt sich das durch den Zusammenhang von Teilchenbewegung und Temperatur erklären. Da die Temperatur ein Maß für die mittlere kinetische Energie eines Stoffs ist, lässt sich verkürzt formulieren: Je höher die Temperatur, desto schneller die Teilchen. Am Phasenübergang zum Gas werden die Teilchen so schnell, dass sie nicht mehr durch die intermolekularen Wechselwirkungen zusammengehalten werden können. Diejenige Flüssigkeit mit dem niedrigeren Siedepunkt geht also in die Gasphase über und lässt sich durch einen geeigneten Apparaturaufbau wieder kondensieren und auffangen. Dabei verbleibt das Gemisch trotz weiterer Wärmezufuhr bei der Siedetemperatur, weil alle Energie in den Phasenwechsel von flüssig zu gasförmig fließt. Die höher siedende Flüssigkeit wird während des Vorgangs so lange flüssig bleiben, bis die niedriger siedende vollständig aus dem Gemisch entfernt wurde. Erst jetzt wird die zugeführte Wärme dafür sorgen, dass sich die Temperatur der verbleibenden Flüssigkeit erhöht und irgendwann selbst zu sieden beginnt. Für die quantifizierende Beschreibung dieses Phänomens bei einem Reinstoff werden Phasendiagramme verwendet, an denen man die entsprechenden Grenztemperaturen und -drücke ablesen kann. Bei Zwei- oder Mehrstoffgemischen kann die Abschätzung komplizierter werden, weil intermolekulare Wechselwirkung einen Einfluss auf die Siedetemperaturen haben können. Dies geht bis hin zur Untrennbarkeit und der Ausbildung eines gemeinsamen Siedepunkts des Gemischs (sog. Azeotrop). Für die in dieser Studie konstruierten Aufgaben ist vor allem der primäre Zusammenhang von Molekularmasse und Siedepunkt

relevant. Umfassendere Beschreibungen von Destillationsprozessen bleiben darum außen vor. Unter der Voraussetzung, dass die zwei Stoffe sich per Destillation trennen lassen, kann die Trennung durch Rektifikation, d. h. Mehrfachdestillation verbessert werden. Hierfür stehen zahlreiche Verfahren zur Verfügung und diese werden auch bei der Erdöldestillation angewendet.¹¹ Obwohl es technisch möglich wäre, lohnt es sich wirtschaftlich nicht jeden einzelnen Reinstoff aus dem Rohöl zu destillieren. Es werden stattdessen so genannte Fraktionen, d. h. Stoffgemische mit einem Siedepunktbereich gewonnen. Tankstellenbenzin besteht beispielsweise nicht aus einem Reinstoff, sondern aus Kohlenwasserstoffen mit einer Siedetemperaturspanne.

Wassermolekül

Die Verbindung Wasser zeichnet sich auf der Stoffebene durch eine Reihe von Besonderheiten aus. Beispielsweise hat es einen, in Relation zu seinem Molekulargewicht $M = 18 \text{ g mol}^{-1}$, extrem hohen Schmelz- und Siedepunkt sowie eine große spezifische Wärmekapazität. Es ist als Transport- und Puffermedium die dominante anorganische Verbindung in Lebewesen und es ist auf der Erdoberfläche nicht nur quantitativ allgegenwärtig, sondern gehört zu den grundlegenden Faktoren globaler klimatischer Zusammenhänge. Gleichzeitig erlaubt der Temperaturbereich für die Phasenübergänge das Vorhandensein aller drei Aggregatzustände auf der Erdoberfläche. Diese Besonderheiten sind auf der submikroskopischen Ebene angelegt. Aufgrund der Elektronegativitätsdifferenz zwischen Wasserstoff und Sauerstoff bilden sich Partialladungen aus, die mit anderen polaren Teilchen in Wechselwirkung treten. Im Vergleich mit Methan mit einem Molekulargewicht $M = 16 \text{ g mol}^{-1}$ bspw., kann die starke Siedepunktdifferenz von ca. 140°C durch die Polarität erklärt werden. Methanmoleküle besitzen durch die unpolare Bindungsverhältnisse eine deutlich niedrigere intermolekulare Anziehung, d. h. es ist deutlich weniger thermische Energie nötig um in den gasförmigen Zustand überzugehen. Mit diesem strukturellen Unterschied ist auch erklärbar, warum Methan sich praktisch nicht in Wasser löst, Methanol mit einem polaren Strukturteil jedoch in jedem Zahlverhältnis mit Wasser mischbar ist. In erster Näherung gilt dabei der Merksatz *Gleiches löst Gleiches*. Mit dem *Gleichen* ist die Polarität gemeint, die durch eine asymmetrische Ladungsverteilung in der Atombindung zustande kommt und so elektrostatische Wechselwirkungen zwischen Molekülen mit ähnlicher Partialladung erzeugt.

Reaktionsgleichungen

In dieser Arbeit werden Reaktionsgleichungen nur im Zusammenhang mit der Nutzung von Summenformeln verwendet und keine Gleichgewichtsreaktionen benutzt, sondern ein vollständiger Stoffumsatz unterstellt. Ladungen spielen ebenfalls keine Rolle. Es werden

¹¹Im Labormaßstab wird für eine einfache Destillation meist der klassische Liebig-Kühler verwendet. Rektifikationskühler (bspw. Vigreux-Kolonnen, oder Kühler mit Raschig-Ringen) können die Trennleistung stark erhöhen. Auf eine tiefere Darstellung wird hier verzichtet, weil auch im Industriemaßstab kein anderes physikalisches Prinzip verwendet wird, sondern die Umsetzung dieses Prinzips rein anlagentechnischer Natur ist.

2. Theoretischer Hintergrund

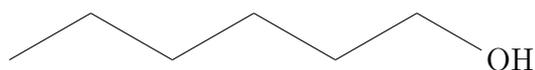


Abbildung 2.2.: So genannte *Skelettförmel* von Hexan-1-ol.

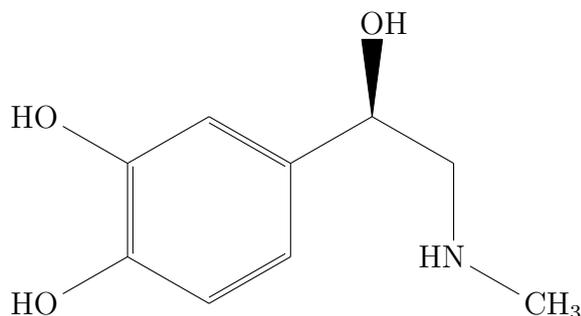


Abbildung 2.3.: Strukturformel von Adrenalin mit räumlicher Auszeichnung der OH-Gruppe.

einer bestimmten Anzahl von Teilchen, deren atomaren Zusammensetzungen, sowie die Richtung der chemischen Reaktion, bei der sich diese Zahlverhältnisse verändern, bestimmte Symbole zugewiesen. Jedes Atom erhält sein entsprechendes Elementsymbol aus dem PSE (bspw. Na), die Anzahl von Molekülen pro Umsatz wird mit Faktoren symbolisiert (bspw. 2Na) und die Anzahl von Atomen innerhalb eines Moleküls mit Indices (bspw. H₂O) angegeben. Der Reaktionspfeil (\longrightarrow) steht für den Prozess der Teilchenumordnung, die von den Edukten ausgeht und zu den Produkten führt. Der Massenerhaltungssatz ist dabei von großer Bedeutung. Er wird, unter Berücksichtigung der Faktoren und Indices, durch die gleiche Anzahl der einzelnen Elementsymbole auf der linken sowie der rechten Seite des Reaktionspfeils zum Ausdruck gebracht:



Strukturformeln

Strukturformeln sind ein Sammelbegriff für eine Vielzahl verschiedener Schreibweisen, deren gemeinsames Merkmal die Darstellung geometrischer Verhältnisse in und zwischen Molekülen ist. Die formale Darstellung ist dabei weniger standardisiert als bei Reaktionsgleichungen mit Summenformeln, weil je nach Verwendungszweck andere Schwerpunkte wichtig sind. Oft sollen organische Moleküle mit minimalem Aufwand gezeigt werden (Abb. 2.2), oder es liegt eine bestimmte räumliche Konfiguration einer funktionellen Gruppe im Fokus (Abb. 2.3). Konventionell werden in der Modellierung von Reaktionsmechanismen die Elektronenverschiebungen mit Pfeilen gekennzeichnet, woraus sich das Brechen und die Neuverknüpfung von Bindungen ergibt. Durch die zusätzliche Angabe von Ladungen und relative Größen der beteiligten Atome, können außerdem elektrostatische und sterische Bedingungen gekennzeichnet werden, die eine Abschätzung von Reaktionsverläufen erlaubt.

2.3.3. Exemplarische Verbindungen der Themen mit modelltheoretischen Aspekten

In allen Bezugsthemen kann der Begriff des Modells nicht einfach wortwörtlich als *modulus* („Maßstab“) verstanden werden. Exemplarisch werden darum die fünf Charakteristika von Modellen (vgl. Abschnitt 2.2.1) konkretisiert, um die Aufgabeninhalte der Arbeit und die Modelltheorie miteinander zu verbinden.

Sowohl Reaktionsgleichungen als auch Strukturformeln sind im Sinne der *repräsentativen Intransparenz* sehr eingeschränkte Darstellungen. Die Abstraktionsleistung bei der Übertragung eines Phänomens auf die Teilchenebene und die formale Beschreibung der dabei ablaufenden Reaktion gehört zum fachsprachlichen Kern der Chemie. Weil Reaktionsgleichungen außerdem auch für hypothetisches Arbeiten verwendet werden, d. h. es möglich ist durch die Zahlverhältnisse der miteinander reagierenden Stoffe experimentelle Vorhersagen und Erklärungen zu generieren, können sie auf der Linie dieser Arbeit als Modelle interpretiert werden. Reaktionsgleichungen sind symbolische Werkzeuge, die in erster Näherung als Abkürzungen verstanden werden können. Anstatt natürliche Sprache eins zu eins in Schriftsprache zu übersetzen, werden bestimmten Zeichen bestimmte Bedeutungen zugewiesen und arithmetische Operationen damit durchgeführt. In diesem repräsentativen Modus sind sie sehr abstrakte Hilfsmittel, weil ein Atomsymbol bspw. für eine definierte Anzahl von Atomen steht.

Der etwas despektierliche Begriff der *Tafelchemie*, also der Darstellung von umfänglichen Reaktionsmechanismen, ohne dass der Chemiker sich in einem Labor befindet, ist ein passendes Beispiel für den hypothetisch-deduktiven Charakter der organischen Synthesechemie. Per Ausschlussprinzip können Annahmen über eine Reaktion und ihren konkreten Verlauf gemacht werden. Dabei geht die theoretische Untersuchung bis zur Übertragung einzelner Elektronen und konstatiert ein klassisch mechanisches Bild von massebehafteten Körpern, die auf bestimmte Arten und Weisen miteinander wechselwirken. Dass in dieser Vorgehensweise fachlich unterkomplex operiert wird,¹² spielt für die Nützlichkeit und Effizienz dieser Vorgehensweise keine primäre Rolle und erfüllt eben die konstruktivistischen Überlegungen zur Modellnutzung, die dieser Arbeit zugrunde liegen.

Für das Nachvollziehen, Erklären oder Planen einer Stofftrennung per Destillation, werden hier beispielhaft zwei Dinge betont. Erstens ist es – zumal beim Lernen im Fach Chemie – schwierig direkt in einen Destillationsturm hineinsehen zu können. Zweitens sind verschiedene Skalierungsüberlegungen notwendig. Wer eine Destillationsapparatur aus dem Labor für die industrielle Rohölraffination einfach vergrößern wollte, müsste zwangsläufig scheitern. Ein filigraner Glaskühler ist ab einer bestimmten Größe viel zu instabil und die abzuführende Wärmemenge nimmt nicht linear, sondern kubisch mit der Vergrößerung des Behälterradius zu. Trotzdem ist das Prinzip (Stofftrennung nach Eigenschaft, hier: der Siedepunkt) in jedem Format gleich. In dieser Hinsicht ist die Abbildung und Beschreibung einer Destillationsapparatur ein Modell für den Destilla-

¹²Damit sind u. a. gemeint: anthropomorphe Zuschreibungen (ein Teilchen *will* ein Elektron *haben*), die Vernachlässigung der Interpretation von Teilchen als Quantenobjekten mit Ortsunschärfe, exakte Ladungs- oder Bindungslokalisierungen bspw. im Benzolring, oder auch starre Stäbe als Bindungen.

2. Theoretischer Hintergrund

tionsturm einer Ö raffinerie. Das Charakteristikum des *beschränkten Designs* spielt hier in der Folge, trotz der zunächst scheinbar rein makroskopischen Dimension sowohl des Modells als auch des realen Gegenstands, eine wichtige Rolle.

Ladungsverschiebungen in Molekülen zu modellieren und zu nutzen ist die Voraussetzung für eine erklärende oder sogar prädiktive Untersuchung von molekularen Wechselwirkungen. Dies gilt sowohl im Allgemeinen, als auch für die Interpretation von Lösungsprozessen mit Wasser als Lösungsmittel im Speziellen. Weil solche Ladungsverschiebung genau genommen in jedem polaren Molekül anders gelagert ist und der Übergang von einer Ionenbindung über eine polare Atombindung hin zu einer unpolaren Atombindung fließend ist, müsste der Merksatz *Gleiches löst Gleiches* eher *Ähnliches löst Ähnliches* lauten. Für das basale Verstehen und Nutzen des Prinzips wird aber hier eine starke Vereinfachung vorgenommen, die sogar in ihrer Fachlichkeit grenzwertig ist: Die Darstellung eines Wassermoleküls als Stabmagnet ist auf den ersten Blick zugänglich, auch weil elektrische und magnetische Felder sich gegenseitig bedingen. Die vereinfachte Darstellung (vgl. Anhang H) kann aber Fehlvorstellungen implizieren, weil nur der Aspekt der Anziehung von negativen und positiven Polen eine Rolle spielt. Weil damit aber bezüglich der *Ergebnisorientierung* ein naturwissenschaftliches Phänomen strukturiert und zur Anwendung gebracht werden kann, ist es modelltheoretisch legitim die Fachlichkeit vorläufig in den Hintergrund zu stellen.

Die *konkrete Manipulierbarkeit* kann am Vergleich von Materialeigenschaften auf der makroskopischen Ebene und ihrer zugrunde liegenden submikroskopischen Struktur illustriert werden. Es ist nicht trivial, den Bruch bzw. das Verformen bestimmter Stoffe auf nicht sichtbare Kugelteilchen mit einer bestimmten Elektronenanordnung zurückzuführen. Unter der hypothetischen Überlegung, ein Stoff *würde* sich auf eine bestimmte Art und Weise verhalten, falls er eine bestimmte Teilchenstruktur *hätte*, kann zwischen verschiedenen Annahmen hin und her gewechselt werden, deren Bewertung nur über die empirische Prüfung abgeschlossen werden kann. Entscheidend ist aber, dass dieses hypothetische Moment die Manipulation des Modellobjekts erlaubt, ohne dass ein Anspruch auf die formale Richtigkeit unter Berücksichtigung allen Fachwissens gestellt werden muss.

Alle genannten Beispiele können in zwei Schwerpunkten zusammengefasst werden. Erstens erlaubt es das Charakteristikum der *pluralen Rechtfertigung* zum Zweck einer naturwissenschaftlichen Untersuchung, bestimmte Fachaspekte vorläufig zu vernachlässigen. Es wäre sogar kontraproduktiv, würde versucht werden stets alles, was über einen Sachverhalt bekannt ist, in eine begrenzte Untersuchung mit einbezogen werden (vgl. Abschnitt 6.1.1). Zweitens ist das hypothetisch-deduktive Vorgehen bei der Modellnutzung im Zusammenspiel mit der Prüfung an Phänomenen oder Daten universell auf Themen im Fach Chemie anwendbar. Dabei spielt die Frage nach der Skala keine Rolle, d. h. sie ist primär unabhängig davon, ob ein Erkenntnisprozess auf der Stoffebene, bei der Untersuchung von einzelnen Teilchen, oder bei der Manipulation von Symbolen abläuft.

2.3.4. Curriculare Einordnungen

Der Übergang von der sichtbaren Stoff- auf die nicht direkt wahrnehmbare Teilchenebene stellt eine große Herausforderung beim Lehren und Lernen im Fach Chemie dar (Barke et al., 2015; Johnstone, 2010; Streller et al., 2019) und der in den Bildungsstandards geforderte Übergang von der Erklärung eines makroskopischen Zustands oder Prozesses zu ihren zugrundeliegenden, submikroskopischen Strukturen ist nicht trivial.

Für Unterrichtsgänge im Fach Chemie ist der Vergleich von Stoffeigenschaften und die Rückführung auf Zusammenhänge auf atomarer Ebene ein basaler Bestandteil. In Berliner Schulen wird das bindend in die jeweiligen Rahmenlehrpläne übersetzt und diese sehen im Kompetenzfeld *Mit Fachwissen umgehen* vier Basiskonzepte vor, von denen zwei (*Stoff-Teilchen-Konzept* und *Struktur-Eigenschafts-Konzept*) direkt für die vorliegenden Themen relevant sind. Es wird unter anderem verlangt, dass SuS „[...] zwischenmolekulare Wechselwirkungen auf Teilchenebene erklären.“ (Landesinstitut für Schule und Medien Berlin-Brandenburg, 2016, S. 17) und dass sie „[...] strukturelle Ordnungsprinzipien von Stoffen (Ionensubstanzen, Molekülsubstanzen, Metalle u. a. begründen.“ (ebd.). Bindungsarten und die Unterscheidung von Stoff- und Teilchenebene werden im Anfangsunterricht Chemie bis zur gymnasialen Oberstufe immer wieder als Grundlage für die Bearbeitung von Phänomenen eingesetzt. Es ist darum davon auszugehen, dass SuS regelmäßig mit Problemstellungen konfrontiert werden, die den Aufgabenbeschreibungen dieser Arbeit ähneln. Weil das beschriebene, technische Verfahren der Destillation auf Prinzipien beruht, die aus dem Teilchenmodell abgeleitet werden können, ist ihre Modellierung in chemischen Lehr-/Lernsituationen sinnvoll und notwendig. Hier wird explizit zwar nicht auf die Teilchen eingegangen, sondern der Zusammenhang von Siedepunkt und Molekularmasse hergestellt, letztere kann aber wiederum auf das Teilchenmodell zurückgeführt werden. Im Chemieunterricht wird die Gewinnung von Erdöl oft mit dem Destillationsverfahren verknüpft, weil es sich gemeinsam mit der homologen Reihe der Alkane und der organischen Synthesechemie in eine sachlogische Abfolge einordnen lässt. Die Destillation als Arbeitstechnik selbst ist kein obligatorischer Inhalt nach dem Berliner Rahmenlehrplan. Die Trennung von Stoffgemischen nach ihren Eigenschaften kann aber im genannten Lehr-/Lernzusammenhang exemplarisch gut bearbeitet werden. Das Wassermolekül ist im Berlin-Brandenburger Rahmenlehrplan der exemplarische Vertreter für Lösungsmittel und den Prinzipien von Lösungsvorgängen. Allein unter Rückbezug auf das Stoff-Teilchen-Konzept, verknüpft dieses Thema die entsprechenden Aufgaben der Studie direkt mit den Vorgaben für den Schulunterricht. Reaktions- und Strukturformeln sind zwar übergreifende, auch fachsprachliche Aspekte im Chemieunterricht, ihre isolierte Thematisierung wird aber vorgenommen, um im empirischen Teil der Arbeit zu prüfen, ob sie aus modelltheoretischer Perspektive von den anderen Themen unterschieden werden müssten. Unabhängig von dieser Forschungsüberlegung, sind sie aber ein unverzichtbarer Teil der Schulchemie, weil sie ein unverzichtbares Kommunikationsmittel des Fachs sind. Insofern wird den Aufgaben der Arbeit curriculare Validität unterstellt (vgl. auch Abschnitt 4.2.1).

2.4. Zusammenfassung

Im vorliegenden Theoriekapitel wurden zunächst ihre epistemologischen Bezugspunkte benannt, dann die Perspektive auf Modelle und ihre Nutzung im Kontext von Chemieunterricht ausgeführt und anschließend mit fachspezifischen Kontexten verknüpft. Es folgt in diesem Abschnitt ein zusammenfassender Überblick, in dem auch die Ziele der Arbeit expliziert werden, um die Forschungsfragen und -hypothesen (Kapitel 3) vorzubereiten. Eine überblicksweises Einordnung der Arbeit in die Forschungslandschaft wird vorgenommen und die Konstruktionsbeschreibung der Aufgaben im Methodenkapitel wird durch die Darstellung der theoretiegeleiteten Struktur konkretisiert.

2.4.1. Theorieüberblick

Die erkenntnistheoretische Grundhaltung dieser Arbeit kann als moderat konstruktivistisch klassifiziert werden (vgl. McCarthy, 2018, S. 119 ff.). Die Existenz von submikroskopischen Teilchen wird nicht in Frage gestellt. Ihre Modellierung seitens der naturwissenschaftlichen Forschung seit John Dalton und die Evidenz der produktiven Nutzung von sich gegenseitig widersprechenden Modellen in den Naturwissenschaften und ihren Didaktiken weisen aber darauf hin, dass der Wahrheitsanspruch für eine Existenz von Atomen in genau der einen oder genau der anderen *Art und Weise* mindestens schwierig zu rechtfertigen ist. Genau das sollte in Lehr-/Lernzusammenhängen berücksichtigt werden. Durch Wissensakkumulation ergibt sich nicht notwendigerweise ein differenziertes Bild über die Vorläufigkeit von naturwissenschaftlichen Erkenntnissen (Zoller & Nahum, 2012) und die Explikation von Denk- und Handlungsstrategien fördert diese nachweislich bei SuS (Brady et al., 2015). Weil gleichzeitig eben solch eine explizite Nutzung von Modellen für hypothetisch-deduktive Untersuchungen im Chemieunterricht selten stattfindet, soll die vorliegende Arbeit einen Beitrag dazu leisten, wie sie als generalisierbares Handlungsmuster einsetzbar gemacht werden könnte. Knuutila (2005, 2011, 2017) dient dabei als Basis, auf der, mithilfe der vier beschriebenen Modellkomponenten (Lesh, 2000) als kommunikative Hilfe, verschiedene Themen so dargestellt werden, dass das modellbasierte, hypothetisch-deduktive Vorgehen in problemhaltigen Umgebungen rekonstruiert werden kann.

2.4.2. Zielkonkretisierung

Durch die Explikation der erkenntnistheoretischen Vorannahmen soll ein Beitrag zur fachdidaktischen Theoriebildung geleistet werden, der chemiespezifisch nicht ausreichend publiziert ist (vgl. Abschnitt 1). Dazu gehört eine tiefere, erkenntnistheoretische Auseinandersetzung (vgl. Abschnitt 2.2.1) mit Modellen und deren Verknüpfung mit einer empirischen Strukturprüfung. Die Verknüpfung zwischen Modellnutzung und hypothetisch-deduktiver Erkenntnisgewinnung soll zwar Bestandteil von Lehr-/Lernprozesse im Fach Chemie sein, sie ist aber empirisch nicht ausreichend charakterisiert (vgl. Abschnitte 2.1 und 2.2. Wenn Modellnutzung eine übergeordnete Fähigkeit innerhalb der Chemie sein soll, sollte sie unabhängig vom gewählten Themeninhalt sein. Die generali-

sierbare Nutzbarkeit des Strukturvorschlags soll daher plausibel gemacht werden (vgl. Abschnitt 2.3). Methodisch soll versucht werden, die Komplexität der konstruierten Aufgaben in passenden, statistischen Modellen abzubilden (vgl. Abschnitt 4.4). Bezogen auf den methodisch-statistischen Teil, sollen die modelltheoretischen Überlegungen zum hypothetisch-deduktiven Vorgehen exemplarisch auch auf die Ergebnisse gewendet werden (vgl. Abschnitt 6).

2.4.3. Einordnung der Arbeit in das Feld der Kompetenzmodellierung

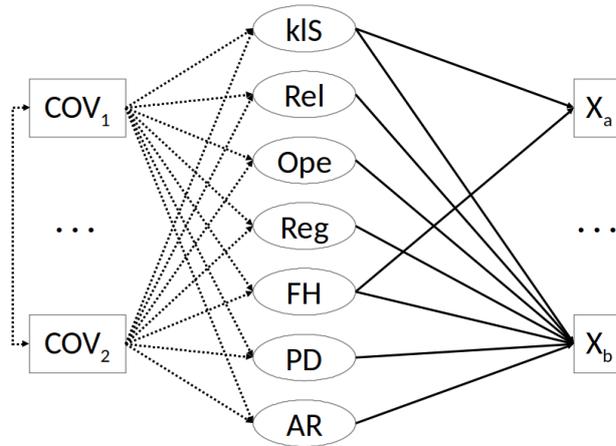
Weil die o. g. Handlungsmuster und die daraus ableitbaren Fähigkeitsfacetten differenziert messbar gemacht werden sollen, kann die Arbeit als Versuch eingeordnet werden, ein deskriptives Kompetenzstrukturmodell zu konstruieren (Schecker & Parchmann, 2006). Parchmann (2010) formuliert die Notwendigkeit der domänenspezifischen Dimensionalitätsprüfung in der Kompetenzmodellierung, die für Modellnutzung in Chemie bisher nicht granular genug (d. h. prozessorientiert über den hypothetisch-deduktiven Dreischritt, vgl. Nehring et al., 2016) oder als stark kognitionspsychologisch geprägte Problemlösestrategie (Scherer & Tiemann, 2014) untersucht ist. Aktuell wird eine quantitative Dimensionsanalyse zu kompetenzorientierter Modellnutzung nur für das Fach Biologie beschrieben (Krüger et al., 2018). Die normativen Vorgaben seitens der Bildungsstandards (vgl. Kapitel 1), erkenntnistheoretische Überlegungen (Abschnitt 2.2.1), sowie fachdidaktische Forschungslücken (s. o.) werden als Anlässe gesehen, eine detailliertere, fachspezifische Beschreibung der Modellnutzung vorzunehmen. In diesem Zusammenhang wird die Fähigkeit zur Modellnutzung nicht als isoliert betrachtet, sondern mit Hilfe eines Hintergrundmodells beschrieben, um die Abhängigkeit von weiteren Konstrukten (bspw. kognitive Fähigkeiten) einbeziehen zu können (vgl. Abschnitt 4.2.3 sowie Carstensen et al., 2007; Frey et al., 2008). Die Operationalisierung der konstruktbezogenen Aufgaben für dieses Vorhaben weicht von den dominant beschriebenen Vorgehensweisen in Lehrbüchern (vgl. Bandalos, 2018; Kauertz, 2014; Moosbrugger & Kelava, 2012b; Neumann, 2014) und der empirisch-fachdidaktischen Praxis (Hartig & Höhler, 2010, S. 191) ab: Üblicherweise werden entsprechende Aufgaben möglichst so konstruiert, dass der Test unidimensional ist, oder eine Aufgabe jeweils *genau einer* Kompetenzdimension zuzuordnen sind (vgl. auch Kauertz et al., 2010). Dieser Ansatz liegt hier nicht vor, weil bspw. in einer Aufgabe zur Hypothesenauswahl mit Hilfe der *Relationen*, bereits mindestens zwei Dimensionen enthalten sind. Zur Bestimmung einer Relation müssen vorher die *kleinsten Sinneinheiten*, zwischen denen diese Relation besteht, identifiziert werden. Zur theoretischen Beschreibung der resultierenden *within-item-dimensionality* (vgl. Abschnitte 2.4.4 und 4.4.2), sowie für den allgemeinen methodisch-statistischen Ansatz Abschnitt 4.4.1. Zur Frage, ob das hier konzipierte Messinstrument trotz seines geschlossenen Formats ein vergleichsweise komplexes Konstrukt abbilden kann, sei auf Abschnitt 4.2.1 verwiesen. Shavelson (2013) kann als kritischer Begleiter für diese Form der Fähigkeitsmodellierung gelesen werden.

2.4.4. Ableitung des zu prüfenden Konstrukts

Die dargestellte Theorie impliziert eine sogenannte *within-item-dimensionality* (Hartig & Höhler, 2008, 2009; Zhang, 2013). Wenn mit Hilfe der Komponente *Operationen* die *Planung und Durchführung* einer naturwissenschaftlichen Untersuchung angesteuert werden soll und in den *Operationen* strukturell auch alle anderen Komponenten enthalten sind, dann folgt daraus für eine entsprechende Aufgabe, dass sie mindestens fünf Dimensionen enthalten sollte.¹³ Im Fall einer Aufgabe, bei der mit den *Regeln* eine *Analyse und Reflexion* durchgeführt werden soll, würde das eine 7-Dimensionalität bedeuten. Gleichzeitig kann die latente Fähigkeitsstruktur kompensatorischer Natur sein (Nehring & Schwichow, 2020; von Davier, 2014), d. h. einzelne Ausprägungen von Teilfähigkeiten können eine geringere Ausprägung einer anderen Teilfähigkeit ausgleichen. Das wäre beispielsweise der Fall, wenn das formale – also nicht inhaltsbezogene – Verstehen von *Wenn-Dann-Beziehungen* in den *Regeln* die grundlegende Teilfähigkeit darstellen und mit einem höherem Gewicht zur Modellnutzung insgesamt beitragen würde. Daher werden die folgenden Modelle vorgeschlagen, die jeweils zur Strukturaufklärung beitragen sollen. Das sogenannte Hintergrundmodell wird für eine bessere Übersicht hier nur als Platzhalter dargestellt. Eine detaillierte Herleitung, Darstellung und Analyse wird in den Abschnitten 5.2 und 5.3 vorgenommen.

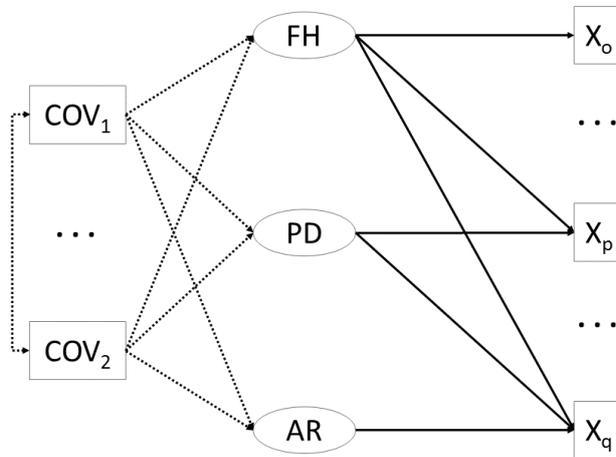
1. Ein 7-dimensionales Modell, das die ineinander geschachtelten Teildimensionen vollständig berücksichtigt. Das bedeutet bspw.: In einer Aufgabe zu den *Relationen* sind automatisch auch *kleinste Sinneinheiten* enthalten, aber nicht umgekehrt. Eine Aufgabe, die sich auf das Identifizieren einer passenden Hypothese bezieht muss keine *Planung und Durchführung* enthalten, umgekehrt aber schon. Abbildung 2.4 zeigt exemplarische Strukturzusammenhänge für je eine Aufgabe zu *Frage/Hypothese* mit *kleinsten Sinneinheiten* und eine Aufgabe zu *Auswertung/Reflexion* mit *Regeln*.
2. Ein 3-dimensionales Modell, das sich allein auf die Teilschritte des hypothetisch-deduktiven Erkenntnisgangs bezieht. Es ist theoretisch plausibel, dass dieses Vorgehen im Umgang mit Modellen dominant ist und die Art der Modellbeschreibung in den Hintergrund tritt. Wer grundsätzlich in der Lage ist, hypothesenbezogen naturwissenschaftliche Untersuchungen vorzunehmen, der muss nicht notwendigerweise mit der detaillierten Struktur des jeweils zugrundeliegenden Modells umgehen können, das so einer Untersuchung zugrunde liegt. Abbildung 2.5 zeigt den Strukturvorschlag.
3. Ein 4-dimensionales Modell, das nur die vier beschriebenen Modellkomponenten für die Aufgabenlösung berücksichtigt. Wenn die Analyse der Teilkomponenten eines Modells vorliegt, dann kann sich allein dadurch schon die Auswahl einer zielführenden Untersuchungsplanung ergeben. Es ist möglicherweise unerheblich, ob

¹³Ohne Hypothese kann per Definition keine Planung durchgeführt werden (vgl. Abschnitt 2.1.1). Die Modellkomponenten sind ähnlich strukturiert. Die Relation zwischen zwei kleinsten Sinneinheiten kann gerade dann identifiziert werden, wenn auf diese Einheiten Bezug genommen wird (vgl. Abschnitt 2.2.2).



Anm.: COV = Kovariable, klS = kl. Sinneinheiten, Rel = Relationen, Ope = Operationen, Reg = Regeln, FH = Frage/Hypothese, PD = Planung/Durchführung, AR = Auswertung/Reflexion, X_a = Aufgabe zu klS & FH, X_b = Aufgabe zu Reg & AR.

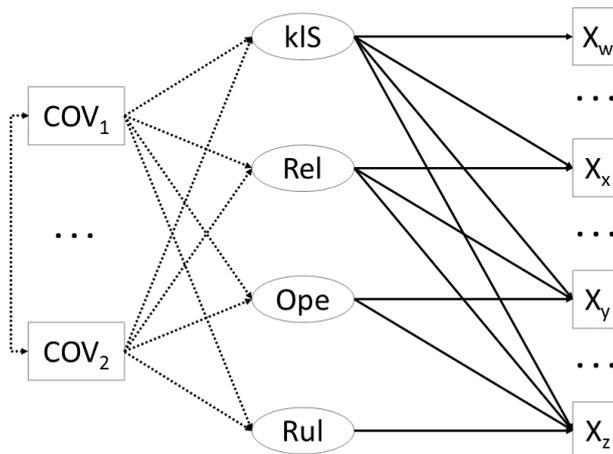
Abbildung 2.4.: Exemplarische Darstellung des 7-dimensionalen Modells.



Anm.: COV = Kovariable, FH = Frage/Hypothese, PD = Planung/Durchführung, AR = Auswertung/Reflexion, X_i = Aufgaben.

Abbildung 2.5.: Exemplarische Darstellung des 3-dimensionalen Modells.

2. Theoretischer Hintergrund

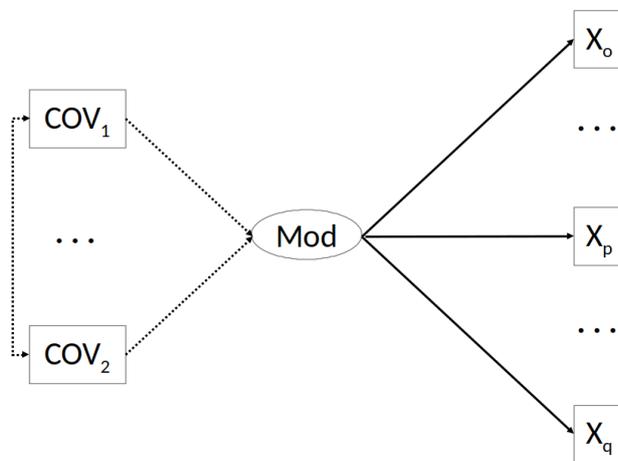


Anm.: COV = Kovariable, kIS = kl. Sinneinheiten, Rel = Relationen, Ope = Operationen, Reg = Regeln, X_i = Aufgaben.

Abbildung 2.6.: Exemplarische Darstellung des 4-dimensionalen Modells.

das hypothetisch-deduktive Vorgehen eine sinnvolle Strategie in den Naturwissenschaften darstellt. Allein die systematische Beschreibung eines Systems könnte die vordergründige Einflussgröße für Erkenntnisgewinnung sein. Abbildung 2.6 zeigt den Strukturvorschlag.

4. Ein 1-dimensionales Modell, das die Modellnutzung in problemhaltigen Umgebungen zufriedenstellend beschreibt. Es muss gar nicht sein, dass die Fähigkeit so differenziert darstellbar ist, wie die Theorie der vorliegenden Arbeit es impliziert. Dass es sich um nur eine dezidierte Fähigkeit handelt ist dabei allerdings theoretisch und empirisch unplausibel. Möglicherweise ist das Nutzen von Modellen für Erkenntnisgewinnung aber ein so vielseitiges Gemisch aus unterschiedlichsten kognitiven Fähigkeitsdimensionen, dass sie der Erfassung mit dem gewählten Vorgehen nicht vollständig zugänglich ist, aber trotzdem als Leistungsmerkmal beschrieben werden kann. Abbildung 2.7 zeigt den Strukturvorschlag.



Anm.: COV = Kovariable, Mod = Modellnutzungsfähigkeit, X_i = Aufgaben.

Abbildung 2.7.: Exemplarische Darstellung des 1-dimensionalen Modells.

3. Forschungsfragen & -hypothesen

Aus der dargestellten Theorie und den vorgeschlagenen Strukturvorschlägen zur Modellnutzung als Mittel der Erkenntnisgewinnung ergeben sich die hier formulierten Forschungsfragen. Ihnen folgen jeweils empirisch zu prüfende Hypothesen, die beibehalten oder zurückgewiesen werden und die die Grundlage für die Beantwortung der Fragen dienen. Die interessierende Forschungshypothese H_1 wird jeweils durch ihre korrespondierende Nullhypothese H_0 ergänzt. Diese wird dann der statistischen Prüfung unterzogen (vgl. Abschnitte 4.4.1 und 5.3). Die aufgeführten Forschungshypothesen basieren auf der hierarchischen Schachtelung der statistischen Modelle, was einen direkten Vergleich verschiedener Dimensionalitätsannahmen erlaubt. Im Zusammenspiel dieser numerischen Argumentation mit den Validitätsüberlegungen (Abschnitt 4.2.1), sowie der dargelegten theoretischen Strukturen (Abschnitt 2.4.4) ergeben sich auch unter Berücksichtigung der Bedingungen der statistischen Prüfungen (Abschnitt 4.4.3) verschiedene Schwerpunktanworten (Abschnitt 5.4, die abschließend gegeneinander abgewogen und diskutiert werden (Kapitel 6).

3.1. Forschungsfrage I

Entspricht die empirische Struktur der Testergebnisse der theoretisch angenommenen Struktur der hypothetisch-deduktiven Modellnutzung im Fach Chemie?

Unter der Annahme, dass die Nutzung von Modellen zum Durchlaufen des hypothetisch-deduktiven Denkens über die vier beschriebenen Modellkomponenten strukturiert werden kann, sollten die konstruierten Testaufgaben, entsprechende statistische Kennwerte zeigen und nicht im Widerspruch zu den theoretischen Überlegungen stehen.

3.1.1. Hypothese Ia

H_1 : *Ein 7-dimensionales Modell passt signifikant besser auf die Testdaten, als ein 4-, 3-, oder 1-dimensionales Modell.*

H_0 : *Höherdimensionale Modelle lassen sich statistisch nicht von Modellen mit geringerer Dimensionalität unterscheiden.*

3. Forschungsfragen & -hypothesen

3.1.2. Hypothese Ib

H_1 : Ein Modell mit einer Within-Item-Dimensionality passt signifikant besser auf die Testdaten, als ein Modell mit einer Einfachstruktur.

H_0 : Ein Modell mit einer Einfachstruktur und ein Modell mit Within-Item-Dimensionality lassen sich statistisch nicht voneinander unterscheiden.

3.2. Forschungsfrage II

Ist die Modellnutzung in der dargestellten Form unabhängig von den jeweiligen Themenbezügen?

Wenn das Nutzen von Modellen in chemiespezifischen Lehr-/Lernumgebungen als verallgemeinerbare Fähigkeit beschrieben werden soll, dann dürften die in den konstruierten Aufgaben verwendeten Themenbezüge keinen Einfluss auf die Lösungswahrscheinlichkeit, d. h. die Aufgabenschwierigkeiten bzw. Personenfähigkeiten haben. Diese Nivellierungsprüfung kehrt das übliche empirisch-statistische Verfahren im Sinne der Forschungsfrage um: Sollte die Nullhypothese nicht zurückgewiesen werden, so wäre die Frage positiv beantwortet.

3.2.1. Hypothese IIa

H_1 : Die mittleren Aufgabenschwierigkeiten aus den IRT-Skalierungen weichen zwischen den Aufgabenthemen signifikant voneinander ab.

H_0 : Zwischen den Aufgabenschwierigkeiten der jeweiligen Themen gibt es keine signifikante Abweichung.

4. Methoden

In diesem Kapitel werden die Methoden zur Beantwortung der Forschungsfragen und der Prüfung der Hypothesen beschrieben. Zu diesem Zweck wird zunächst ein Überblick der Planung und Durchführung des empirischen Studienteils gegeben. Der Kern des Projekts ist dabei die quantitative Hauptstudie, die mit Hilfe entsprechender Voruntersuchungen konzipiert wurde. Das iterativ erarbeitete Konstruktionsmanual der Testaufgaben, der Datenmanagementplan, sowie die verwendeten statistischen Erfassungs- und Auswertungsmethoden werden dargelegt. Auf diese Ausführungen wird im Ergebnisteil (Kapitel 5) entsprechend verwiesen.

4.1. Der Studienplan im Überblick

Für die Bearbeitung empirisch-fachdidaktischer Studien sehen Schecker et al. (2014) verschiedene Formate vor, die die Gesamtanlage des Vorhabens charakterisieren (ebd., S. 4). Hier sind die *Bezugsdisziplinen* die Erkenntnis- und Wissenschaftstheorie innerhalb des Fachs Chemie (vgl. Abschnitt 2.2.1), das *Erkenntnisinteresse* liegt bei der Strukturaufklärung eines kognitiven Prozesses (hier: Modellnutzung für hypothetisch-deduktives Denken, vgl. Abschnitt 2.4) und in der *methodischen Grundausrichtung* ist die Arbeit als quantitative Querschnittsstudie mit einem schriftlichen Multiple-Choice-Leistungstest als zentralem Instrument angelegt (vgl. Döring & Bortz, 2016c, 2016d). Da die Erhebung in Kooperation mit Schulen stattfand, deren Aufwand minimiert werden sollte, stand bereits bei frühen Projektskizzen fest, dass die SuS in so genannten *Klumpenstichproben* (vgl. Bortz & Schuster, 2010, S. 79 ff.) an der Testung teilnehmen würden.

Für die Planung und Durchführung dieser Art von Studie lassen sich Mustervorgehensweisen zur Untersuchung von kognitiven Dimensionen sozialwissenschaftlicher Untersuchungen im Allgemeinen (vgl. bspw. Bandalos, 2018; Jonkisz et al., 2012) oder auch für naturwissenschaftsdidaktische Projekte (Riese & Reinhold, 2014) mit SuS als Zielgruppe im Speziellen (vgl. Terzer et al., 2013) identifizieren. Das auf die vorliegenden Forschungsfragen (vgl. Abschnitt 3) adaptierte Vorgehen ist in Abbildung 4.1 illustriert. Nach der ersten Aufarbeitung des Literaturstandes (Rost & Tiemann, 2017) wurden unterschiedliche Fachinhalte (Abschnitt 2.3) ausgewählt und Modellbeschreibungen konstruiert, die zur Beschreibung des jeweiligen Systems dienen sollten. Ein Expertenrating (Abschnitt 5.1.1) zur Zuordnung und Identifizierung der Modellkomponenten (Abschnitt 2.2.2) in den einzelnen Themen sicherte die fachlich-inhaltliche Kohärenz und Konsistenz ab. Es folgte eine erste Vorstudie mit einem qualitativen Schwerpunkt: SuS der 9. Jahrgangsstufe eines Gymnasiums wurden mit offenen Aufgaben zur Identifizie-

4. Methoden

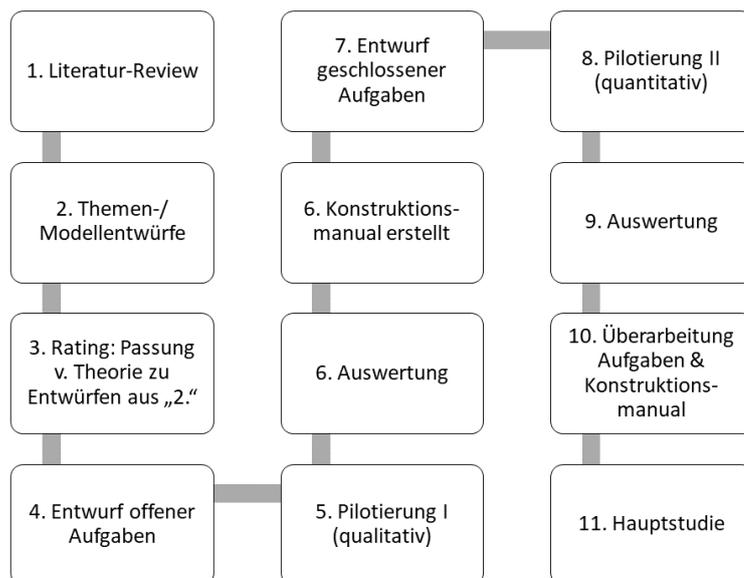


Abbildung 4.1.: Schematische Darstellung der Datenerhebung.

Die Bewertung von Modellkomponenten aus gegebenen Modellen aufgefordert (Abschnitt 5.1.2). Die Antworten wurden im Rahmen einer Masterarbeit ausgewertet und flossen in die Konzeption und die Überarbeitungsschritte des Konstruktionsmanuals (Abschnitt 4.2.2) für die geschlossenen Aufgaben ein. Die zweite Vorstudie (Abschnitt 5.2) war quantitativ angelegt. An zwei Gymnasien und einer Integrierten Sekundarschule wurden die insgesamt 60 Aufgaben (Abschnitt 4.2.2) inklusive standardisierter Tests für kognitive Fähigkeiten, Leseverstehen und weiterer Kovariablen (Abschnitt 4.2.3) von SuS der 10. Jahrgangsstufe bearbeitet. Die erhaltenen Ergebnisse wurden für eine weitere Überarbeitung des Konstruktionsmanuals und in der Folge zur Überarbeitung der Aufgaben (bspw. Reformulierung der Distraktoren) verwendet. Nach der Entscheidung, die Hauptstudie (Abschnitt 5.3) auf die 10. Jahrgangsstufe von Gymnasien einzuschränken, wurden die Aufgaben im Rahmen einer weiteren Masterarbeit final überprüft und homogenisiert, bspw. durch Prüfung auf Wortzahl des Aufgabenstamms und Fachwortdichte. Mit der Durchführung der Hauptstudie wurde die Datenerhebung des Projekts abgeschlossen.

4.2. Die Konstruktion des Erhebungsinstruments

Bezüglich methodischer Vorgehensweisen in der fachdidaktischen Forschung stehen eine Reihe von Befunden zur Verfügung (vgl. Krüger et al., 2014). Induktiv gewonnene Strukturvorschläge aus eher qualitativen Ansätzen zur Modellnutzung (Dass et al., 2015) können dabei natürlicherweise keinen Anspruch auf Generalisierbarkeit erheben. Die Ergebnisse aus stärker quantitativen Vorhaben stellen sich trotz eines stärkeren Fokus auf Verallgemeinerungen als schwer vergleichbar heraus (vgl. Kapitel 1). Dies wird in der vorliegenden Arbeit als Folge fehlender (Scherer, 2014) oder nur unzureichender (Stieff et

al., 2016) Explikation der erkenntnistheoretischen Grundannahmen gedeutet. Da bereits theoretische Ansätze konstruiert worden sind, deren Für und Wider mit vergleichsweise geringeren Stichprobenumfängen abgewogen wurde, sollte der nächste Schritt ein strukturaufklärendes Verfahren sein. Da die angenommene Struktur (vgl. Abschnitt 2.4.4) relativ komplex ist, bedurfte es eines ökonomischen Vorgehens, das über multiple-choice Aufgaben realisiert wurde (Bühner, 2011; Tiemann & Körbs, 2014).

4.2.1. Testgütekriterien quantitativer Erhebungen mit multiple-choice Aufgaben

Schriftliche Leistungstests unterliegen einer Reihe von Bedingungen, wenn verallgemeinerbare Populationsaussagen aus ihnen generiert werden sollen. Die drei Hauptgütekriterien *Objektivität*, *Reliabilität* und *Validität* werden folgend überblicksweise skizziert und durch – für diese Studie relevante – *Nebengütekriterien* ergänzt. In der Fachliteratur findet sich eine Vielzahl von Quellen für Testgütekriterien. Für diese Studie wird vor allem auf Moosbrugger und Kelava (2012b) sowie Bortz und Döring (2016) Bezug genommen. Die dargestellten Maßnahmen zur bestmöglichen Beachtung der Kriterien beziehen sich vordergründig auf die Hauptstudie, weil die Vorstudien auch zur Erfassung und Dokumentation von regelmäßig auftretenden Phänomenen wie Störungen oder Nachfragen von SuS dienten und planungsgemäß weniger standardisiert angelegt waren. Die entsprechenden Erkenntnisse und die daraus abgeleiteten Handlungsalternativen flossen dann in die Überlegungen für die Haupterhebung ein (vgl. Abschnitte 5.1 und 5.2).

Objektivität

Nach Moosbrugger und Kelava (2012a) ist ein Test dann objektiv,

„[...] wenn er dasjenige Merkmal, das er misst, unabhängig von Testleiter und Testauswerter misst. Außerdem müssen klare und anwenderunabhängige Regeln für die Ergebnisinterpretation vorliegen.“ (ebd., S. 8)

Es sei zielführend in die drei Teilbereiche der *Durchführungs-*, *Auswertungs-* und *Interpretationsobjektivität* zu unterteilen.

Um größtmögliche Durchführungsobjektivität zu gewährleisten, wurde ein Instruktionsskript erstellt, das für den Haupttestleiter an verschiedenen Terminen einerseits und für unterstützende Testleiter andererseits eine standardisierte Wortwahl für verschiedene Testabschnitte vorgab. Darin enthalten war auch eine Auflistung von verschiedenen, aus den Vorstudien abgeleiteten, Fragen, die die SuS regelmäßig stellten. Beispielsweise sah das Instruktionsskript vor, den SuS zu erklären, dass sie jeweils 15 Fragen zu Modellen beantworten sollten. Regelmäßig wurde die noch nicht abgeschlossene Erklärung mit der Frage, ob man alle Fragen auf einmal beantworten könne, oder ob man jeweils eine bestimmte Zeit abwarten solle unterbrochen. Auch die Hilfe bei der Aufgabenbearbeitung durch teils anwesende Lehrkräfte wurde in den Vorstudien beobachtet und musste unterbunden werden. Für die unterstützenden Testleiter wurde eine Schulung durchgeführt,

4. Methoden

die, mit dem Instruktionsskript als Leitfaden, solche und andere Herausforderungen vorweg nahm und ein einheitliches Vorgehen zur Folge hatte.

Weil in der Hauptstudie ausschließlich ein geschlossenes Aufgabenformat mit a priori festgelegter Richtig-Falsch-Dichotomie vorlag, war nicht anzunehmen, dass Einschränkungen der Auswertungsobjektivität vorliegen würden.

Die Interpretationsobjektivität für die standardisierten Instrumente zweier Kovariablen (*LGVT* und *KFT*, vgl. Abschnitt 4.2.3) war durch die entsprechenden Normwerttabellen und die dokumentierte Verlässlichkeit der Instrumente sichergestellt. Für die objektive Fähigkeitsschätzung des Fachwissenstests wird ebenfalls auf Abschnitt 4.2.3, sowie Abschnitt 5.2.2 verwiesen. Das Zusammenspiel der Maßnahmen zur Modellgeltungsprüfung (vgl. Abschnitt 4.4.3) soll verschiedene Perspektiven in der Ergebnisdeutung ermöglichen. So wird dieses Kriterium auch im engen Zusammenspiel mit der Validität (s. u.) gesehen.

Reliabilität

Die kurze Definition der Reliabilität ist das, was der Test misst, (möglichst) ohne Messfehler zu messen. Diese idealisierte Annahme ist prinzipiell nicht erreichbar, weswegen die Reliabilität als numerischer Kennwert in verschiedenen Berechnungsansätzen stets Werte zwischen Null und Eins – mit der Eins für perfekt reliable Messinstrumente – annimmt (vgl. Moosbrugger & Kelava, 2012a; Schermelleh-Engel & Werner, 2012). Im vorliegenden Projekt werden hauptsächlich statistische Verfahren aus der Familie der *Item-Response-Theorie* (IRT) eingesetzt, deren Reliabilitätsmaße von denen aus der so genannten *klassischen Testtheorie* (KTT) abweichen. Das dominante Maß wird dabei der *Expected-A-Posteriori*-Schätzer (EAP) sein, der als Reliabilitätsangabe für IRT-basierte Tests dokumentiert ist (Bond & Fox, 2015; Wu et al., 2016). Eine exakte Beschreibung der verwendeten Reliabilitätsschätzungen wird in Abschnitt 4.4.1 vorgenommen weil sie unmittelbar in die Methodenbeschreibung eingebunden ist. Die Interpretation als Kennwert für die Messgenauigkeit ist aber für beide Paradigmen vergleichbar: Was immer der Test misst, misst er mehr oder weniger präzise. Dabei macht die Reliabilität keine Aussage darüber ob das, was gemessen wird, tatsächlich erfasst wird (s. u.).

Validität

Wie im Abschnitt über die Reliabilität angedeutet, wird mit der Validität in erster Näherung angegeben, ob ein Test das angezielte Merkmal tatsächlich misst. Nach der Standardliteratur ist diese einfache Beschreibung als zu verkürzt zu beurteilen. Sowohl Hartig et al. (2012, S. 144 ff.), als auch Bortz und Döring (2016, S. 93 ff.) betonen in ihren angegebenen Definitionen das *begründete Urteil* über die Validität eines Tests über empirische Befunde und theoretische Überlegungen, die der Inferenz unterliegt und nicht automatisch durch eine Methode oder ein Forschungsdesign herbeigeführt werden kann. Sie kann aber dem übergeordneten Kriterium der *methodischen Strenge* zugeordnet werden (ebd.). Insofern die Grundfrage der vorliegenden Arbeit (*Misst das konstruierte Instrument die Modellnutzung im Rahmen des hypothetisch-deduktiven Denkens?*)

eng damit zusammenhängt, werden folgend verschiedene Perspektiven zur Validitätsprüfung dargestellt, um sie im Diskussionsteil (Abschnitt 6) wieder aufzugreifen und in die Überlegungen einzubeziehen. Falls nicht anders angegeben, werden die entsprechenden Abschnitte aus den beiden o. g. Quellen herangezogen. Weitere Überblicksarbeiten bieten Field et al. (2012) sowie Schermelleh-Engel et al. (2006).

Der Begriff der *Konstruktvalidität* bezieht sich auf die Interpretation der Messergebnisse durch die theoriebasierten Annahmen. Dazu gehört vor allem die a-priori-Definition der Zusammenhänge zwischen den einzelnen Facetten des Konstrukts untereinander (hier: Modellkomponenten), sowie mit Kovariablen. Diese Zusammenhänge bilden das *nomologische Netz*, dessen Knotenpunkte mit *Korrespondenzregeln*, d. h. Korrelations- oder Wirkrichtungsannahmen, miteinander verbunden sind (siehe bspw. *KFT* in Abschnitt 4.2.3, oder Dimensionalitätshypothesen aus Abschnitt 3). Die Bedeutung der *Konstruktvalidität* bei fachdidaktischen Projekten führen Schmiemann und Lücken (2014) aus. Die Untersuchung der Dimensionalität kann unter dem Begriff der *faktoriellen Validität* verhandelt werden. Diese wird mit statistischen Methoden entweder falsifiziert oder vorläufig beibehalten, jedoch prinzipiell *nicht* bestätigt. Die *statistische* Validität wird durch die Einhaltung der Voraussetzungen der jeweiligen Methoden (vgl. Abschnitt 4.4.1) sichergestellt. Sie unterliegt aber bereits vorab Beschränkungen, weil bestimmte Kennwerte zur Qualitätsbestimmung der Aufgaben und des Gesamttests (bspw. bestimmte Fit-Indices) nur durch einen vollständigen Datensatz berechnet werden können. Dieser wird per Design (vgl. Abschnitt 4.2.4) nicht vorliegen. Insofern kein randomisiertes Experiment unter strenger Variablenkontrolle vorliegt, ist das Kriterium der *internen* Validität nach KTT zunächst nicht vollumfänglich erfüllbar. Es handelt es sich bei der Untersuchung also formal um eine Feldstudie (Bortz & Döring, 2016). Bei IRT-Modellen kann die interne Validität allerdings durch die hinreichend dargestellte Modellgeltung abgesichert werden (Rost, 2006). Für eine zusätzliche Absicherung gegen Zufallseffekte, werden die theoriegeleiteten Strukturmodelle mit solchen Modellen verglichen, bei denen die Aufgaben zufällig den jeweiligen Dimensionen zugeordnet werden (vgl. Abschnitt 4.4.1). Weil die SuS in ihrer gewohnten Umgebung arbeiten und mit Hilfe der o. g. Maßnahmen die *Objektivität* möglichst hoch gehalten wird, kann das Kriterium der *externen* Validität aus konvergenter und diskriminanter Perspektive bearbeitet werden. Nach dem konvergenten Ansatz soll gezeigt werden, wie die Ergebnisse mit Studien ähnlicher Konstrukte zusammenhängen. Das diskriminante Vorgehen soll die Konstruktvalidität ex negativo stützen, indem bspw. deskriptiv ein nicht zu hoher Zusammenhang von Lesefähigkeit und Lösungswahrscheinlichkeit der konstruktbezogene Aufgaben gezeigt würde. Schließlich soll die *Inhaltsvalidität* durch die enge Verbindung von Theorie und Aufgabenkonstruktion gewährleistet, sowie *curriculare* Validität durch Voruntersuchungen und Abgleich mit Lehrplänen sichergestellt werden.

Nebengütekriterien

Obwohl *Ökonomie*, *Zumutbarkeit* und *Nützlichkeit* zwar nicht so dominant beschrieben werden, wie die Hauptgütekriterien (vgl. Bortz & Döring, 2016; Moosbrugger & Kelava, 2012b), spielen sie in der konkreten Umsetzung eines schulbezogenen Forschungspro-

4. Methoden

jekts eine wichtige Rolle. Unterrichts- bzw. aktive Lernzeit ist für Lehrkräfte – sowohl objektiv-institutionell als auch empfunden – eine so wertvolle Ressource, dass es bei der Planung und Durchführung einer Befragung einer engen Abstimmung und sensibler Kommunikation bedarf. Die Realisierung einer genügend großen Stichprobe, die gleichzeitig außerhalb des Unterrichts und in genügend kontrollierter Umgebung stattfindet, war im Rahmen der zur Verfügung stehenden Ressourcen seitens der Forscher nicht zu gewährleisten. Das wäre auch im Sinne der *Zumutbarkeit* nur schwer zu rechtfertigen gewesen: In einem schulbezogenen Forschungsvorhaben sollte eine Störung des Alltags der SuS so gering wie möglich sein, um die Balance aus Erkenntnisinteresse (d. h. auch der *Nützlichkeit*, vgl. Abschnitt 2.4) und Probandenbeeinträchtigung möglichst moderat zu gestalten. In diesem Zusammenhang sei auf Abschnitt 4.2.4 verwiesen, dass das Testheftdesign beschreibt. Es diene dazu, eine höchstmögliche Bearbeitungszahl pro Aufgabe bei gleichzeitig kleinstmöglicher Probandenbelastung zu gewährleisten.

4.2.2. Aufgabenkonstruktion

Aus der theoretischen Rahmung der Studie ergeben sich mit der Auswahl von fünf Themen (vgl. Abschnitt 2.3), drei Schritten des hypothetisch-deduktiven Vorgehens (Abschnitt 2.1.1) und der vier Modellkomponenten (Abschnitt 2.2.2) formal 60 Aufgaben, wenn jede Kombination in jeweils einer Aufgabe ausgedrückt werden soll (vgl. Abschnitt 2.4.4). Für das Erstellen von Leistungstestaufgaben ist es notwendig, das Konstrukt, das beibehalten oder verworfen werden soll, so exakt wie möglich abzubilden. Vermeidbare Einflussgrößen sollten minimiert oder ganz ausgeschaltet werden, nicht vermeidbare Einflussgrößen sollten durch zusätzliche Messungen kontrolliert werden (vgl. Ropohl et al., 2015; Walpuski & Ropohl, 2014, siehe auch Abschnitt 4.2.3). Das zu messende Konstrukt sollte dann über alle Aufgaben hinweg systematisch abgebildet werden (Mayer & Wellnitz, 2014). Zu diesem Zweck wurde ein Konstruktionsmanual konzipiert, das als Leitfaden für die Aufgabenerstellung diene. Es stellte eine formale Gleichheit und eine inhaltliche Vergleichbarkeit sicher, indem es für jede Zelle der theoretischen Struktur (vgl. Abschnitt 2.4.4) Bausteine definierte. Abbildung 4.2 zeigt ein Beispiel aus dem Manual und Abbildung 4.3 ein Aufgabenbeispiel. Das vollständige Manual und weitere Aufgabenbeispiele sind in den Anhängen F und H hinterlegt. Auf diese Weise wurden die allgemeineren Anleitungen (Bandalos, 2018; Jonkisz et al., 2012) und die spezielleren Umsetzungen (Nehring, 2014; Terzer et al., 2013) der systematischen Aufgabenkonstruktion auf die vorliegende Studie angepasst.

4.2. Die Konstruktion des Erhebungsinstruments

Modellzelle: FH_re	Fragestellungen und Hypothese zu Merkmalen und Zusammenhängen mit Hilfe der Relationen von Modellen entwickeln.
Modellbestandteil: Relationen	Denkschritt: Fragestellung und Hypothese
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell
Frage	Welche Vermutung können sie mit Hilfe des Modells aufstellen?
Antwortmöglichkeiten	Eine Hypothese die widerspruchsfrei auf Basis des Itemstamms und der Relationen des Modells abgeleitet werden kann. Drei Hypothesen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Relationen des Modells abgeleitet werden können.
gemessene Fähigkeit	Zuordnung einer Hypothese zu gegebenen Informationen auf Basis der Relationen eines chemiespezifischen Modells.

Abbildung 4.2.: Konstruktionsschema für eine Aufgabe zu *Frage/Hypothese* mit *Relationen* beim Thema *Destillation*.

DES_FH_re

Hintergrundinformationen

Erdöl besteht aus verschiedenen Kohlenwasserstoffen und Verunreinigungen. In einer Ö Raffinerie werden aus diesem Stoffgemisch verschiedene *Fraktionen* in verschiedenen Mengen gewonnen, die unterschiedliche Zwecke erfüllen und verschiedene Summenformeln besitzen (Tabelle 1). Dies geschieht in einem so genannten Destillationsturm, in dem die Fraktionen voneinander getrennt werden.

Abbildung 1 zeigt eine Glasapparatur, die als Modell für eine Ö raffinerie dient.

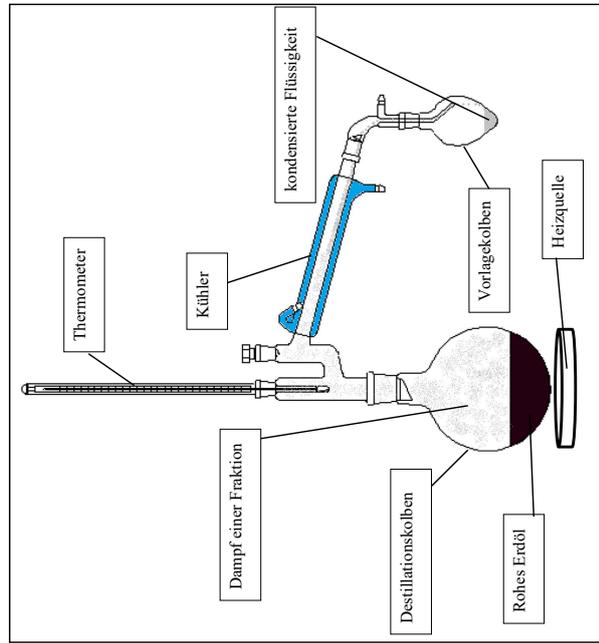


Abbildung 1: Eine Destillationsapparatur als Modell für einen Destillationsturm.

Im Destillationskolben wird rohes Erdöl erhitzt. Die Fraktion, die bei der niedrigsten Temperatur siedet, steigt gasförmig auf und gelangt am Thermometer vorbei zum Kühler. Dort kondensiert das Gas zur Flüssigkeit und tropft in den so genannten Vorlagekolben. Anschließend siedet die nächste Fraktion. Das erkennt man an einem plötzlichen Anstieg der Temperatur am Thermometer. Wechselt man während des Temperaturanstiegs den Vorlagekolben, so kann man die Fraktionen gut voneinander trennen.

DES_FH_re

Tabelle 1: Siedetemperaturen und Verwendungen verschiedener Fraktionen im Erdöl.

Siedetemperatur T in °C	Bezeichnung der Fraktion	Summenformel	mögliche Verwendungen
20	Gas, Flüssiggas	CH ₄	Heizgas, Kunststoffproduktion
100	Leichtbenzin	C ₄ H ₁₀	Treibstoffe (Benzin an der Tankstelle)
180	Schwerbenzin	C ₆ H ₁₄	Treibstoffe für spezielle Motoren, Kunststoffproduktion
240	Kerosin	C ₁₀ H ₁₈	Flugzeugtreibstoff
360	Gasöl	C ₁₄ H ₂₆	Treibstoff für Dieselmotoren, Öl Heizten

Aufgabe

Diana und Yannik sind Chemielaboranten in Ausbildung und wollen das Prinzip eines Destillationsturms besser verstehen. Sie meinen, dass es dafür wichtig ist auf bestimmte **Verhältnisse zwischen den Einzelteilen des Modells** zu achten. Das Verhältnis ist in diesem Fall der **Unterschied in den Siedetemperaturen zwischen zwei Fraktionen**.

Welche Vermutung können sie mit Hilfe der Modellapparatur aufstellen? **Kreuze an!**

Je geringer ihr Siedetemperaturunterschied ist, desto ...

- ... ähnlicher verhalten sich zwei Fraktionen.
- ... weniger Verunreinigungen finden sich in zwei Fraktionen.
- ... schwerer lassen sich zwei Fraktionen voneinander trennen.
- ... schlechter lassen sich zwei Fraktionen kondensieren.

Innerhalb jedes Themas beziehen sich die Aufgaben aufeinander. Stehen bspw. die Relationen zwischen den kleinsten Sinneinheiten in einer Aufgabe im Fokus, dann stehen genau diese Sinneinheiten in der korrespondierenden Aufgabe im Fokus. Das gilt ebenfalls für die Dimension des hypothetisch-deduktiven Denkens: Die Auswertung/Reflexion in einer Aufgabe bezieht sich auf eine Hypothese in der korrespondierenden Aufgabe des entsprechenden Kontexts. Das hat den Vorteil, dass für jeden Kontext ein konsequentes, hypothetisch-deduktives Vorgehen konstruiert werden konnte, das sich systematisch auf die Nutzung der Modellkomponenten bezieht. Der messtheoretische Nachteil ist eine starke Abhängigkeit (vgl. Annahme lokaler stochastischer Unabhängigkeit für Rasch-Skalierungen in Abschnitt 4.4.1) zwischen den Aufgaben, der mit einem entsprechend komplexen Testheftdesign (vgl. Abschnitt 4.2.4) begegnet werden musste. Außerdem wurden die Aufgaben so konstruiert, dass alle notwendigen Sachinformationen für die korrekte Schlussfolgerung gegeben wurden. Das führt dazu, dass alle Attraktoren und Distraktoren der Aufgaben grundsätzlich aus den gegebenen Inhalten deduziert werden können. Der jeweilige Attraktor ist dann so konstruiert, dass er *widerspruchsfrei* geschlussfolgert und damit als richtige Antwort klassifiziert werden kann.

4.2.3. Darstellung und Begründung der Kovariablenauswahl

In quantitativen Projekten zur Prüfung von – theoretisch angenommenen – latenten Fähigkeiten sollte sichergestellt werden, dass die gemessene Performanz (bspw. Rohpunktwerte) möglichst ohne den Einfluss bereits validierter und messbarer Konstrukte (bspw. Intelligenz) zustande kommt. Da es gleichzeitig unplausibel ist, dass die Performanz frei von solchen Einflüssen ist, wurden die geschätzten Fähigkeitsparameter der *Lesefähigkeit*, der *allgemeinen kognitiven Fähigkeiten* und des *aufgabenspezifischen Fachwissens* (s. u.) in ein Hintergrundmodell (vgl. Abschnitt 2.4) aufgenommen.

Weil die Bearbeitung der Aufgaben grundsätzlich nicht ohne das Lesen des relativ komplexen Aufgabenstamms (vgl. Beispielaufgaben in Abschnitt 4.2.2 oder Anhang H) möglich ist, ist ein Einfluss durch das Leseverständnis zu erwarten. Weil die individuellen Lesefähigkeiten der Probanden nicht a priori geprüft werden konnten, war es angemessen diese Einflussgröße zu kontrollieren. Die Messung erfolgte über den *Lesegeschwindigkeits- und Verständnistest* (LGVT) nach Schneider et al. (2007).

Für allgemeine kognitive Fähigkeiten stehen in der Psychologie zahlreiche Messverfahren zur Verfügung, die sich in der Regel aus dem Zweifaktor-Modell der Intelligenz und dessen Substrukturen ableiten lassen (Vock & Holling, 2006). Für die Untersuchung fachbezogener kognitiver Facetten ist es notwendig, den Einfluss solcher übergeordneten Konstrukte zu kennen und sie davon zu unterscheiden. Sonst wäre nur schwer zu rechtfertigen, warum die Lösung der konzipierten Aufgaben der spezifisch angenommenen Fähigkeit nicht einfach mit der Formel *fluide Intelligenz plus Fachwissen* zu erklären wäre. In diesem Fall hinge die Lösungswahrscheinlichkeit der Aufgaben in erster Näherung allein davon ab, spezifische Fachinformationen erfolgreich miteinander vernetzen zu können. Dass das *auch* der Fall ist, ist plausibel. In welchem Maß das der Fall ist, muss empirisch prüfbar gemacht werden. In fachdidaktischen Forschungsprojekten ist die Skala zu *figuralen Analogien* des *Kognitiven Fähigkeitstests* (KFT) nach Heller und

4. Methoden

Perleth (2000) ein häufig eingesetzter Vermittler von Intelligenz. Für eine Begründung wird auf Baumert et al. (2009) verwiesen, die neben der Diskussion über das Verhältnis von Intelligenz und Large-Scale-Assessments wie bspw. PISA, die genannte KFT-Skala als verlässliche Schätzmethode für den G-Faktor der Intelligenz darstellen (ebd., S. 170).

Da das Ziel der Arbeit ist, eine verallgemeinerbare aber fachspezifische Fähigkeitsbeschreibung vorzunehmen, stellen auch individuelle Fachwissensunterschiede eine Einflussgröße dar. Das ist auch im Sinne des Erhebungsdesigns zielführend. Die SuS sind naturgemäß in Schulen und diese wiederum in Klassen mit teils wechselnden Lehrpersonen als sogenannte Klumpenstichproben aggregiert. Die Annahme homogener Fachwissensstände ist also schon auf Basis institutioneller Rahmenbedingungen nicht haltbar. Weil darüber hinaus auch in fachlich dekontextualisierten Studien (Bühner et al., 2008) das Inhaltswissen zu problemorientierten Aufgaben einen nachweisbaren Einfluss hat, ist eine Schätzung über die spezifischen Kenntnisse zu Aufgabeninhalten in chemiebezogenen Aufgaben eine notwendige Ergänzung des Testinstruments. Tabelle 4.1 listet die konstruktrelevanten Kovariablen der vorliegenden Arbeit auf. Weitere Variablen, die

Tabelle 4.1.: Konstruktrelevante Kovariablen.

Variable	Instrument (Abkürzung)	Quelle
allg. kognitive Fähigkeiten	Kognitiver Fähigkeitstest (KFT)	Heller und Perleth (2000)
Leseverständnis	Lesegeschwindigkeits- und Verständnistest (LGVT)	Schneider, Schlagmüller und Ennemoser (2007)
Fachwissen	Fachwissenstest (FW)	selbst konstruiert

Anm.: –

für deskriptive Zusammenhänge erhoben wurden waren die Deutsch-, Mathematik- und Chemienoten, das Alter, das Geschlecht, sowie Datum und Uhrzeit der Erhebung. Mit den Noten wird eine externe Fähigkeitsschätzung der SuS in die Studie einbezogen. Die Untersuchung der Fähigkeit zur Modellnutzung sollte ins Verhältnis mit Leistungskriterien im Schulalltag gesetzt werden, um zumindest approximativ anschlussfähig zu realen Unterrichtsprozessen zu sein (vgl. Abschnitt 5.2.2). Alters- und geschlechtsbezogene Unterschiedsprüfungen wurden erhoben um ein vollständigeres Bild der SuS aus der Stichprobe zu erhalten. Der Einfluss des Geschlechts auf die Lösungswahrscheinlichkeit ist für naturwissenschaftliche Leistungen nachweisbar (OECD, 2019b; Schipolowski et al., 2019) und gleichzeitig ist eine entsprechende Begriffs- und Theoriebildung bei weitem noch nicht vollständig für naturwissenschaftsdidaktische Forschung ausdifferenziert (für einen Überblick vgl. Scantlebury, 2012). Zeitliche Angaben zur Erhebung wurden als notwendig für die Einordnung von als unzuverlässig bewerteten Fällen (siehe Abschnitt 4.3) antizipiert.

4.2.4. Testheftdesign

In der Studie sollten 60 Aufgaben eingesetzt werden (vgl. Abschnitt 4.2.2). Für die unter Abschnitt 4.2.1 beschriebene Minimierung der Probandenbelastung und die Vermeidung von Ermüdungseffekten (Hartig & Buchholz, 2012) während der Testung, konnten nicht alle Aufgaben von allen Teilnehmern bearbeitet werden. Die Abschätzung der Bearbeitungszeit pro Aufgabe aus der Vorstudie und die Berücksichtigung der zu erfassenden Kovariablen ergab eine Aufgabenanzahl von 15 Modellaufgaben pro Testheft. Darüber hinaus waren die Aufgaben per Konstruktion voneinander abhängig, weswegen eine sorgfältige Trennung über verschiedene Testheftversionen realisiert werden musste. Für quantitative Multiple-Choice-Studiendesigns schlagen Frey et al. (2009) das Prinzip des *incomplete block designs* (ICBD) vor. Dabei werden Aufgaben in Gruppen zusammengefasst, die über verschiedene Testheftversionen und in verschiedenen Reihenfolgen (sog. *rotiertes Testheftdesign*) so miteinander kombiniert werden, dass für Reihenfolgeeffekte (Wu, 2010) kontrolliert werden kann und die Verknüpfung der Testhefte über mehrfach verwendete Aufgaben sichergestellt wird. Es folgt die Beschreibung der Anpassungen auf Grundlage der Bedingungen für einen Spezialfall des ICBD, nämlich des *balanced incomplete block designs* (BIBD). Dieses diente als Vorlage für das tatsächliche Design, musste aber angepasst werden. An den entsprechenden Stellen werden schematische Abbildungen für das Design gezeigt. Das vollständige Testheftdesign der Studie liegt in Anhang G vor.

1. Jeder Aufgabenblock ist höchstens einmal pro Testheft vertreten. Dieses Kriterium wurde erfüllt.
2. Jedes Testheft hat die gleiche Länge (d. h. Anzahl Aufgaben) und enthält die gleiche Anzahl an Aufgabenblöcken. Dieses Kriterium wurde mit Einschränkungen erfüllt.
3. Jedes Aufgabenblockpaar erscheint über die Testhefte gleich häufig. Dieses Kriterium wurde mit Einschränkungen erfüllt.
4. Jeder Aufgabenblock erscheint gleich häufig über die Testhefte. Dieses Kriterium wurde nicht erfüllt.

Das Testheftdesign sollte über die gesamte Studie für zwei mögliche Merkmalseinflüsse durch die Aufgaben kontrollieren. Es kann einerseits angenommen werden, dass jemand, allein durch die formale Gleichförmigkeit, Aufgaben am Ende eines Testhefts mit höherer Wahrscheinlichkeit löst, wenn er ausschließlich nach einer bestimmten Modellkomponente oder einem bestimmten Denkschritt gefragt wird (Lerneffekt). Andererseits kann angenommen werden, dass Aufgaben am Ende eines Testhefts systematisch schwerer werden (Ermüdungseffekt). Hartig und Buchholz (2012) demonstrieren solche systematischen Itempositionseffekte in den PISA-Daten. Um beide Effekte in der vorliegenden Studie tatsächlich messen zu können und die Ergebnisse entsprechend zu korrigieren, wären vergleichsweise viele Probanden nötig. Da das Erkenntnisinteresse der Studie aber nicht auf der statistischen Modellierung solcher Positionseffekte liegt, wurde

4. Methoden

per Testheftdesign lediglich versucht diese Einflüsse zu minimieren. Aus diesem Grund sollte, erstens, innerhalb eines der fünf Kontexte grundsätzlich kein Testheft sowohl eine Aufgabe zur Planung/Durchführung, als auch eine korrespondierende Aufgabe zur Auswertung/Reflexion vorliegen (vgl. Abschnitt 4.2.2). Das gilt auch für alle weiteren Kombinationen innerhalb dieser Dimension. Zweitens sollte die Reihenfolge der Modellkomponenten und der Denkschritte systematisch variiert werden, um einen Trainings-/Ermüdungseffekt, der die Lösungswahrscheinlichkeit hätte beeinflussen können, über die Gesamtstichprobe zu nivellieren. Weil pro Heft 15 Aufgaben eingesetzt werden sollten, mussten für die Variation nach Denkschritten (drei Stück) und Modellkomponenten (vier Stück) unterschiedlich lange Aufgabenblöcke konstruiert werden. Diese Überlegungen führten zur Konstruktion von drei Testhefttypen gleicher Aufgabenlänge, die die o. g. Bedingungen für ein klassisches BIBD nicht vollständig einhalten konnten.

1. Testhefttyp *Random*

Es wurden drei Testhefte konstruiert, die Aufgaben in zufälliger Reihenfolge enthielten und die in je drei Blöcke mit je fünf Aufgaben unterteilt wurden. Jeder Block tauchte je einmal zu Beginn, in der Mitte und am Ende jedes Testheftes auf. So ergab sich eine Gesamtzahl von neun Heften dieses Testhefttyps.

2. Testhefttyp *Denkschritt*

Es wurden sechs Testhefte á fünf Aufgabenblöcke konstruiert. Jeder Block enthielt nur Aufgaben eines bestimmten Denkschritts, sodass jeder Denkschritt zweimal an jeder Blockposition stand.

3. Testhefttyp *Modellkomponente*

Um die vier Komponenten systematisch über die Testhefte zu verteilen, mussten Aufgabenblöcke mit drei, sowie mit vier Aufgaben realisiert werden, die jeweils nur Aufgaben bezüglich einer bestimmten Modellkomponente enthielten. Auf diese Weise stand – mit leichten Positionsverschiebungen – jede Komponente dreimal am Beginn, dreimal in der Mitte und dreimal am Ende eines Testhefts. Das ergab insgesamt zwölf Testhefte dieses Typs.

Die erhaltenen Testhefte wurden spiralisiert, d. h. in stets fortlaufender Reihenfolge klassen- und schulübergreifend (vgl. Frey et al., 2009) ausgegeben. Es ergaben sich für die Hauptstudie Testhefte, die jeweils insgesamt 80 Minuten Bearbeitungszeit benötigten. Tabelle 4.2 fasst die enthaltenen Fragen/Variablen zusammen.

4.3. Durchführung der Erhebung

Der spiralisierte Einsatz der Papier-und-Stift-Testhefte wurde, wie oben beschrieben, teils von verschiedenen, geschulten Testleitern durchgeführt. Zur Testdurchführung gehörte, neben der präzisen Verwendung des Durchführungsskripts, auch die Dokumentation von Beobachtungen während der Testung. (Moosbrugger, 2012) empfiehlt Fallauschlüsse bei der Datenanalyse aufgrund eines „inadäquaten Bearbeitungsstils“ (ebd., S.

Tabelle 4.2.: Überblick des Erhebungsinstruments.

Reihenfolge	Fragebogen
1.	LGVT
2.	KFT
3.	pers. Daten (Alter, Geschlecht, Noten)
4.	Fachwissen
5.	Modellaufgaben

Anm.: –

246). Dazu wurden bei jeder Erhebung Notizen angefertigt, um das entsprechende Testheft beim Einsammeln zu markieren und später im Datensatz mit einer dichotomen *Vorsicht*-Variable zu versehen. Beobachtungsbeispiele waren u. A.: fremdsprachliche Austauschschüler, Testabbruch wegen außerschulischen Terminen, oder auch obszöne Zeichnungen über das gesamte Testheft hinweg. In Abschnitt 4.3.2 ist der allgemeine Umgang mit Datenausfällen beschrieben und Abschnitt 5 zeigt die, auf die jeweilige Teilstudie bezogenen Entscheidungen über Fallausschlüsse.

Die ausgefüllten Testhefte wurden klassenweise in eine vorbereitete Maske in der Tabellenkalkulationssoftware *MS Excel* übertragen und so mehrere Rohdatensätze erzeugt. Ein Datenmanagementplan (vgl. Abschnitt 4.3.1, s. u.) sicherte die nachvollziehbare Speicherung, Zusammenführung und weiterführende Verarbeitung bzw. Analyse.

4.3.1. Datenmanagementplan

Der Umgang mit Messdaten wird natürlicherweise auf der Basis des jeweiligen Erkenntnisinteresses (vgl. Kapitel 3 geleitet. Darüber hinaus ist es aber hilfreich, auch abstraktere Vorgehensweisen in die Verarbeitungs- und Analysestrategien einzubeziehen. Der Bereich der *Data Science* (DS) erfährt hierfür zunehmende Bedeutung, weil dort allgemeingültige und reproduzierbare Prinzipien beschrieben sind, die die Sicherheit und Nachvollziehbarkeit erhöhen (Hui, 2019; Peng & Matsui, 2016; Sauer, 2019). Die vorliegende Arbeit hat die statistische Modellierung einer Fähigkeit zum Ziel. Aus diesem Grund wurde ein (für diese Arbeit angepasstes) Vorgehen nach Wickham und Grolemund (n. d.) verwendet, das sich explizit als hypothesenorientiert versteht. Im Kontrast zu anderen Zweigen der DS (bspw. *Unsupervised Machine Learning*, bei dem rein datenbasiert explorative Muster erzeugt werden), besteht insofern eine hinreichende Passung für quantitative fachdidaktische Projekte mit verhältnismäßig eingeschränkter Datenbasis.

1. *Datenbereinigung und -aufbereitung*: Gegeben sei der gesamte Rohdatensatz, inklusive noch nicht ausgeschlossener Fälle und fehlender Werte. Nach dem Urteil über auszuschließende Fälle wurden diese entfernt, Rohwerte (bspw. aus dem KFT) in Richtig-Falsch-Dichotomie übersetzt und ggf. fehlende Werte aus den Kovariablen imputiert (bspw. Abschnitte 4.3.2 & 5.3.4).

4. Methoden

2. *Explorative Darstellung*: Mit einer explorativen Datendarstellung von Rohpunktsummen und deren Visualisierung wurde ein Überblick des Datensatzes generiert (bspw. in Abschnitt 5.2).
3. *Parameterschätzungen und statistische Modellierungen*: Nach der Datenübersicht sind, mit Hilfe statistischer Methoden (vgl. Abschnitt 4.4), entsprechende Modellierungen vorgenommen worden. Der Fokus liegt auf Methoden der IRT, weil sie für Untersuchungen der Konstruktvalidität, bzw. Dimensionalität als besser geeignet bewertet werden (Moosbrugger, 2012). Die erhaltenen, statistischen Modelle bilden so die Grundlage für das Beibehalten oder Verwerfen der angenommenen Struktur des theoretischen Konstrukts.
4. *Ergebnisdarstellung*: Für eine übersichtliche Präsentation der erhaltenen Kennwerte und Schätzungen lässt sich nicht in jedem Fall ein standardisiertes Vorgehen ausmachen. Tabellen werden bei aggregierbaren Zahlen (bspw. Spannweite der Fit-Werte von IRT-Modellen) verwendet, für Gruppenvergleiche sind bspw. Boxplots geeignet. Die Arbeit geht nach den Vorgaben der *American Psychological Association* (APA) vor und verwendet das entsprechende Manual (APA, 2013).

Verwendete Software

Mit Ausnahme der Übertragung der Paper-und-Stift-Tests in elektronische Form, wurden alle beschriebenen Schritte der Datenverarbeitung, inklusive der statistischen Methoden (vgl. Abschnitt 4.4) mit der freien Software *R* (R Core Team, 2020) umgesetzt. Diese Software enthält Basisfunktionen (bspw. für lineare Regressionen), sowie eine Vielzahl spezialisierter Ergänzungen – so genannte *Pakete* – für diverse Anwendungen. Die jeweils verwendeten Pakete sind in Tabelle 4.3 aufgelistet und per Angabe der Autor/innen im Literaturverzeichnis hinterlegt.

Datenschutz

Das gesamte Projekt unterliegt den Datenschutzbestimmungen des Landes Berlin für Erhebungen an Schulen¹, sowie den Grundsätzen zum Umgang mit Forschungsdaten der Humboldt-Universität zu Berlin.² Eine Veröffentlichung der Rohdaten, wie durch die Forschungsdatengrundsätze empfohlen, findet trotz einer Anonymisierung nicht statt, weil dies per Genehmigung des Forschungsprojekts explizit ausgeschlossen werden musste. In der Folge sind für die vorliegende und ggf. weitere Publikationen allein die aggregierten bzw. modellierten Daten und Parameterschätzungen zulässig. Das einzige eindeutige Unterscheidungsmerkmal der SuS in den Daten ist eine fortlaufende Nummer, die keine Rückschlüsse auf die zugehörigen Personen zulässt.

Die Datenverarbeitung und -sicherung findet ausschließlich auf zwei lokalen Arbeitsrechnern innerhalb der Universität, einem für solche Zwecke eingerichteten Server, sowie

¹<https://www.forschungsdaten-bildung.de/genehmigungen?la=de>, letzte Einsicht: 18.02.2020.

²<https://www.cms.hu-berlin.de/de/dl/dataman/infos/policy/guidelines/guidelines>, letzte Einsicht: 18.02.2020.

Tabelle 4.3.: Verwendete R-Pakete.

Verarbeitungsprozess	Paketname	Version	Autor-/innen
Basisprogramm	<i>r-base</i>	3.6.3	R Core Team (2020)
Datenbereinigung, -aufbereitung & -visualisierungen	<i>tidyverse</i> & <i>broom</i>	1.3.0 0.7.0	Wickham et al. (2019) Hayes und Couch (2020)
Transfer von Tabellen in das Schreibprogramm	<i>knitr</i>	1.29	Xie (2020)
erweiterte Datenvisualisierungen	<i>ggpubr</i> & <i>statsExpressions</i> & <i>ggstatsplot</i>	0.4.0	Kassambara (2020)
Levene-Test auf Varianzhomogenität	<i>car</i>	3.0-4	Fox und Weisberg (2019)
deskriptive Datenanalyse	<i>psych</i>	1.9.12	Revelle (2019)
IRT-Skalierungen	<i>TAM</i> & <i>mirt</i>	3.5-19 1.32.1	Robitzsch et al. (2020) Chalmers (2012)
Dunn-Test für post-hoc- Gruppenvergleiche	<i>FSA</i>	0.8.30	Ogle et al. (2020)
Effektstärke (ϵ^2) für Gruppenvergleiche	<i>rcompanion</i>	2.3.25	Mangiafico (2020)

Anm.: Versionsnummern sind aufgrund laufender Updates nicht notwendigerweise diejenigen, mit denen tatsächlich die jeweiligen Ergebnisse produziert wurden.

der universitätsinternen Cloud-Lösung (*HU-Box*) statt. Alle Systeme sind passwortgeschützt, Zugang haben nur zwei Personen (Autor & Arbeitsgruppenleiter).

4.3.2. Umgang mit fehlenden Werten

Bei empirischen Studien im Bereich der Sozialwissenschaften ist das Auftreten fehlender Werte (*Missings*) die Regel, nicht die Ausnahme (Döring & Bortz, 2016b; Pohl et al., 2013). In einem schulbezogenen Forschungsprojekt sollte schon vor der Datenerhebung eine Strategie zum Umgang mit Missings erstellt werden, um den erwarteten Ausfällen systematisch begegnen zu können. In der vorliegenden Arbeit wurde dazu die übliche Einteilung in drei Kategorien herangezogen (Spieß, 2010, S. 117 ff.). Die Basis für die Unterstellung dieser drei Ausfallmechanismen ist das Urteil darüber, ob die Wahrscheinlichkeit für das Fehlen eines Wertes abhängig von einem anderen *beobachteten* und *interessierenden* Wert ist. Der allgemeine Umgang mit diesen Mechanismen wird folgend beschrieben, die konkrete Darstellung und die abgeleiteten Konsequenzen (bspw. die nicht designbedingte Missing-Quote oder Fallausschlüsse auf Grundlage von Erhebungsbeobachtungen) erfolgt im Ergebnisteil (Kapitel 5).

Missing-Completely-At-Random (*MCAR*)

Als *MCAR* bezeichnet man fehlende Werte, deren Auftreten durch keine systematische Wechselwirkung mit weiteren interessierenden Variablen erklärt werden können. Das bedeutet: Die Wahrscheinlichkeit, dass der entsprechende Wert fehlt, ist unabhängig von jedweder interessierender Variable, egal ob sie beobachtet wurde oder nicht. Flüchtigkeitsfehler gehören bspw. in diese Kategorie. *MCARs* werden allgemein als unproblematisch betrachtet, weil sie keinen verzerrenden Einfluss auf die Datenanalyse haben, bzw. weil entsprechende Missings durch plausible Werte ersetzt werden können (Spieß, 2010). Lütke et al. (2007) empfehlen dazu sowohl (multiple) imputations-, als auch modellbasierte Verfahren.

Missing-At-Random (*MAR*)

MARs können auch als weniger problematisch eingestuft werden. Das Urteil über ihren Einfluss bedarf aber einer genaueren Untersuchung, weil sie nicht vollständig zufällig zustande kommen. Ein fehlender Wert einer interessierenden Variable (bspw. bei einer Aufgabe zur Modellnutzung) in Abhängigkeit einer anderen interessierenden, aber beobachteten Variable (bspw. Zugehörigkeit zu einer Schule), ist ein Beispiel für diesen Ausfallmechanismus. Weil für die Abhängigkeit des Datenausfalls von einer anderen beobachteten Variable kontrolliert werden kann, ist es möglich sie ohne Verzerrungen zu kompensieren. Die *Missings-By-Design*, die durch das Testheftdesign der Studie entstehen, können insofern als zufällige Ausfälle interpretiert werden, obwohl sie systematisch vorliegen (Frey et al., 2009).

Missing-Not-At-Random (MNAR)

Der *MNAR*-Mechanismus liegt vor, wenn der Ausfall einer interessierenden Variable von der Variable selbst abhängt. Wenn eine Aufgabe so schwer ist, dass die Fähigkeit, die mit dieser Aufgabe gemessen werden soll, nicht für eine Lösung ausreicht, dann liegt *MNAR* vor. Der Umgang mit diesem Mechanismus ist als kritisch zu bewerten, weil keine systematische Korrektur vorgenommen werden kann. Aus diesem Grund sollte er möglichst vermieden werden.

4.4. Statistische Methoden

Folgend werden die statistischen Methoden beschrieben, die im Ergebnisteil für die Analysen verwendet werden (vgl. Kapitel 5) und im Diskussionsteil (vgl. Kapitel 6) interpretiert werden. Methoden aus der IRT (vgl. Abschnitt 4.4.1) stehen dabei im Vordergrund. Die IRT wird überblicksweise dargestellt, Kennwerte für Modellvalidierungen (vgl. interne Validität durch Modellgeltung in Abschnitt 4.2.1) erläutert und die Nutzung multidimensionaler IRT-Modelle (Abschnitt 4.4.2) an Beispielstudien aus dem Feld, sowie am Theorieteil (vgl. Abschnitt 2.4.4) begründet. Den Abschluss bilden Modellvergleichsmaße, die ein Urteil darüber zulassen, ob die theoretischen Dimensionalitätsannahmen empirisch gestützt werden können, oder ob sie verworfen werden müssen (vgl. Kapitel 3).

4.4.1. Grundlagen der Item Response Theorie

Als *Item Response Theory* (IRT) bezeichnet man eine Familie statistischer Methoden, die auch als *probabilistische Testtheorie* bezeichnet wird (vgl. Bühner, 2011, S. 478 ff.). Die Grundlage der IRT – bei Richtig-Falsch-Aufgaben, wie sie in dieser Studie vorliegen – ist die Annahme, dass es eine von Null verschiedene Lösungswahrscheinlichkeit der Leistungsaufgaben gibt. Diese Lösungswahrscheinlichkeit ist zunächst abhängig von der Personenfähigkeit (d. h. des angenommenen kognitiven Konstrukts, bzw. seiner jeweiligen Subdimensionen) und der Aufgabenschwierigkeit, die als äquivalent interpretiert werden: Bei konstanter Personenfähigkeit ist die Lösungswahrscheinlichkeit für eine Aufgabe umgekehrt proportional zu ihrer Schwierigkeit und bei konstanter Aufgabenschwierigkeit steigt die Lösungswahrscheinlichkeit mit steigender Personenfähigkeit. Die messtheoretische Grundannahme ist also, dass die Personenfähigkeit und die Aufgabenschwierigkeit auf der gleichen Skala abgebildet werden. Um die Fähigkeitsschätzung einer Person angeben zu können bedarf es einer mathematischen Operation, die von Richtig-Falsch-Zählungen (respektive: Vergabe eines Punkts bei korrekter Aufgabenlösung) zu einer Wahrscheinlichkeitsangabe führt. Dies nennt man ein Messmodell und es erlaubt Fähigkeitsschätzungen sowie Performanzvorhersagen über die erfassten Rohpunktzahlen hinaus. Die konzeptuelle Grundlage dafür ist das so genannte *Ein-Parameter-Logistische-(1-PL)-Modell*. Der Einsatz dieses Modells ist in der naturwissenschaftlich-fachdidaktischen Forschung weit verbreitet (Boone & Scantlebury, 2006; Hartig & Frey, 2013; Kauertz, 2014; Moosbrugger, 2012; Neumann, 2014) und wird durch eine Reihe von darauf aufbauenden Modellen erweitert, deren Nutzung vor allem bei der Untersuchung komple-

4. Methoden

xerer Konstrukte empfohlen wird (Harvey, 2016). Aufgrund seiner Wichtigkeit wird folgend das 1-PL-Modell ausführlicher beschrieben als die Erweiterungen, für die jeweils entsprechende Quellen angegeben sind. Es wird erweitert durch das *Zwei-Parameter-Logistische*- (2-PL)-Modell, weil es bei der Dimensionalitätsuntersuchung eine wichtige Rolle spielt. Die Steigungsparameter können nämlich als Äquivalente der Faktoren in faktoranalytischen Verfahren der KTT gedeutet werden (Chalmers, 2012; Paek et al., 2018) und erlauben insofern konfirmatorische Zuordnungen von Aufgaben zu theoretisch hergeleiteten Fähigkeitsdimensionen. Das Schätzverfahren für die entsprechenden Parameter ist ein Maximum-Likelihood-Ansatz (Bandalos, 2018; Bock & Aitkin, 1981; Wu et al., 2016), der in der verwendeten Software (vgl. Abschnitt 4.3.1) implementiert ist.

Das 1-PL-Modell

Das 1-PL-Modell verknüpft die Fähigkeit Θ einer Person mit der Aufgabenschwierigkeit β , indem die Rohpunktwerte der Aufgaben zweimal transformiert werden. In einem ersten Schritt wird der Wettquotient der jeweiligen Aufgabe bestimmt. Dieser Quotient ist das Ergebnis aus der deskriptiven Wahrscheinlichkeit die Aufgabe zu lösen ($p_{x=1}$), geteilt durch die korrespondierende Gegenwahrscheinlichkeit ($p_{x=0} = 1 - p_{x=1}$). Der zweite Schritt, die Logarithmierung, führt zum *Logit* der Lösungswahrscheinlichkeit p für eine bestimmte Person v bei einer bestimmten Aufgabe i (Gleichung 4.1):

$$\ln \frac{p(x_{vi} = 1)}{p(x_{vi} = 0)} = \Theta_v - \beta_i \quad (4.1)$$

Umgestellt nach der Lösungswahrscheinlichkeit, erhält man Gleichung 4.2:

$$p(x_{vi} = 1) = \frac{e^{(\Theta_v - \beta_i)}}{1 + e^{(\Theta_v - \beta_i)}} \quad (4.2)$$

Bezogen auf eine konkrete Aufgabenantwort einer bestimmten Person, ergibt sich Gleichung 4.3:

$$p(x_{vi} = x) = \frac{e^{[x_{vi}(\Theta_v - \beta_i)]}}{1 + e^{(\Theta_v - \beta_i)}} \text{ mit } x \in \{0, 1\} \quad (4.3)$$

In Folge der Logarithmierung liegt der Fähigkeits-/ Schwierigkeitsbereich zwischen $-\infty$ und $+\infty$, was eine wichtige Interpretationshilfe ist. Auch eine sehr schwierige Aufgabe hat bei einer Person mit niedriger Fähigkeitsausprägung eine von Null verschiedene Lösungswahrscheinlichkeit. In der Praxis wird dabei meist eine Skala mit Werten zwischen -3 und $+3$ mit der 0 als Mittelwert erhalten, was wie eine z-Standardisierung zu lesen ist. Die Personenfähigkeit und die Aufgabenschwierigkeit sind additiv miteinander verknüpft und führen zu einer leicht zu deutenden Lösungswahrscheinlichkeit: Eine hohe Fähigkeit wird mit einer hohen Aufgabenschwierigkeit kompensiert und umgekehrt. Eine Fähigkeitsausprägung $\Theta = 0$ resultiert bei einer Aufgabenschwierigkeit $\beta = 0$ dann zu einer Lösungswahrscheinlichkeit $p = 0,5$. Für eine umfangreichere Herleitung vgl. Rost (2006) sowie Bühner (2011).

Das 2-PL-Modell

Erweitert man das 1-PL-Modell um einen aufgabenspezifischen Trennschärfeparameter α_i , so erhält man ein Maß dafür, wie gut eine Aufgabe zwischen fähigeren und weniger fähigeren Personen bei $p = 0,5$ unterscheidet (Gleichung 4.4):

$$p(x_{vi} = x) = \frac{e^{[x_{vi}\alpha_i(\Theta_v - \beta_i)]}}{1 + e^{[\alpha_i(\Theta_v - \beta_i)]}} \text{ mit } x \in \{0, 1\} \quad (4.4)$$

Unter der Annahme von Eindimensionalität des zu erfassenden Konstrukts würde man über alle Aufgaben eine relativ homogene Trennschärfe erwarten, wenn sie eine homogene Reliabilität aufweisen. Für Tests mit nur moderaten Reliabilitäten bieten 2-PL-Modelle bessere Parameterschätzungen (Robitzsch et al., 2017; Stewart, 2012), was einen pragmatischen Grund für ihre Nutzung darstellt, weil es die Fähigkeitsmessung und deren Prädiktionskraft präzisiert. Aber „[i]n der psychologischen Testpraxis ist es weitaus häufiger der Fall, dass *unterschiedliche* Trennschärfen durch Itemmehrdimensionalität resultieren.“ (Bühner, 2011, S. 506). Aus diesem Grund gehört die Nutzung des 2-PL-Modells sowohl im explorativen Teil der Analyse (Abschnitt 5.2) als auch bei der konfirmatorischen Modellschätzung der Hauptstudie (Abschnitt 5.3) zu den zentralen Werkzeugen dieser Studie.

Parameterschätzungen

Um Populationsparameter (hier: latente Fähigkeiten) aus Stichproben zu ermitteln, werden im Rahmen dieser Studie sogenannte *Maximum-Likelihood*-(ML)-Methoden angewendet (s. o.). Die Grundfrage dieses Ansatzes ist stets, wie hoch die Plausibilität (*Likelihood*) ist, dass Stichprobenmesswerte (Richtig-Falsch-Antworten auf Testfragen) erhalten werden – gegeben, dass eine Annahme über die den Messwerten zugrunde liegende Größe (Verteilung, Lage und Struktur einer latenten Fähigkeit) zutrifft. In der verwendeten Software (Abschnitt 4.3.1) sind ML-Methoden implementiert, die auf dem Satz von Bayes (Döring & Bortz, 2016a, S. 615 ff.) beruhen und damit Verteilungsannahmen (z. B. Standardnormalverteilung) über Fähigkeiten voraussetzen (vgl. Chalmers, 2012; Fox, 2010; Wu et al., 2016). Der Vorteil dieser Methoden ist eine präzise Parameterschätzung auch bei vergleichsweise vielen Missings, sofern es dabei um Populationswerte geht (Wu et al., 2016, S. 272). Nachteile sind – unter sonst gleichen Bedingungen – eine vergleichsweise geringe Präzision bei der Individualschätzung (ebd.), die Notwendigkeit zur Auswahl eines geeigneten *Priors*, d. h. der begründeten Verteilungsannahme *vor* der eigentlichen Schätzung (Fox, 2010), sowie das Fehlen verallgemeinerbarer Fit-Statistiken zur Modellgeltungsprüfung (Levy, 2011). Vor allem die gute Eignung bei Populationsuntersuchungen führt dazu, dass diese Methoden trotz ihrer Nachteile weit verbreitet sind. Sie liefern stabile Ergebnisse bei Simulationsstudien (Chalmers et al., 2017) und werden in der Analyse von Schulleistungsstudien eingesetzt (Kauertz, 2014; Neumann, 2014; Robitzsch et al., 2017). Im Hinblick auf die Ausrichtung der Arbeit (vgl. Abschnitt 2.4.3) wird die Nutzung dieser ML-Methoden aus diesen Gründen als zielführend bewertet.

Für die Fähigkeitschätzungen werden zwei Maße verwendet. Die *Expected-A-Posteriori*-

4. Methoden

Plausible-Values (EAP-PV) für Populationsfähigkeiten (Bandalos, 2018; Wu, 2005) und die *Weighted-Maximum-Likelihood-Estimates* (WLE) für Individualfähigkeiten (Warm, 1989; Wu et al., 2016). Erstere sind nicht nur für Sekundäranalysen und Korrelationsstudien nützlich (von Davier et al., 2009), sie können auch für die Konstruktion von Fit-Indices verwendet werden (Chalmers & Ng, 2017) und liefern so einen Beitrag zur Modellgeltungsprüfung (vgl. Abschnitt 4.4.3). Letztere sind, wie oben beschrieben, unter den gegebenen Untersuchungsbedingungen zwar weniger reliabel, sie erlauben aber ebenfalls approximative Modellgeltungsprüfungen durch die Abschätzung von systematischen Modell- (Robitzsch et al., 2017) und Gruppenunterschieden (Nehring, 2014).

Frey et al. (2008) sowie Carstensen et al. (2007) beschreiben darüber hinaus die Nutzung von Hintergrundmodellen für die Parameterschätzung von Fähigkeitstests. Dieses Vorgehen ist auch in der pädagogisch-psychologischen Forschung (vgl. beispielhaft Baumert et al., 2009), sowie speziell für Anwendungen in der IRT Paek und Cole (2020)) etabliert und erlaubt die Untersuchung von Testergebnissen mithilfe eines latenten Regressionsverfahrens. Dabei werden die Antwortmuster unter der Berücksichtigung von Kovariablen miteinander ins Verhältnis gesetzt und so ihr Einfluss auf die Testleistung direkt quantifizierbar gemacht.

Zusammenhänge mit Kovariablen

Mit den genannten EAP-PVs werden in den quantitativen Vorarbeiten dieses Projekts Variablenzusammenhänge zwischen geschätzten Modellnutzungsfähigkeiten und Kovariablen vorgenommen (Abschnitt 5.2.2). Dazu werden pro Studienprobandin mehrere zufällige Werte aus der Verteilung der gesamten Populationsschätzung gezogen und mit den so erhaltenen Datensätzen jeweils statistische Zusammenhänge mit den Kovariablen hergestellt. Deren Mittelwerte erlauben im Anschluss stark verzerrungsgeminderte Trendabschätzungen (Laukaityte & Wiberg, 2017; von Davier et al., 2009; von Davier & Hastedt, 2009; Wu, 2005) und erfolgen für die Vorstudie modellbasiert per multipler Imputation (Rubin, 1987). In der Hauptstudie werden keine EAP-PVs mit den Kovariablen korreliert, sondern diese direkt per latenter Regression in die Strukturprüfungen integriert (Hartmann et al., 2015; Paek & Cole, 2020; Scherer, 2011).

4.4.2. Multidimensionale Item Response Theorie

Kompetenzorientierte Aufgaben in den Naturwissenschaften sind in den meisten Fällen als inhärent multidimensional anzunehmen (Ackerman et al., 2003). Selbst unter Konstanthaltung von Kovariablen (bspw. Lesefähigkeit), sollte bei den verwendeten Aufgaben, aufgrund ihrer Struktur (Abschnitt 2.4.4) ein dimensionsprüfendes Verfahren zum Einsatz kommen. Dazu sind explorative und konfirmatorische Verfahren bei IRT-Modellen vorhanden, die auf der bedingten Kovarianz aller Aufgabenpaare basiert und die sich in der *DETECT*-Statistik zusammenfassen lassen (Bonifay et al., 2015; Jang & Roussos, 2007; Seo et al., 2017; Zhang, 2013). Tabelle 4.4 zeigt die Orientierungswerte nach (Bonifay et al., 2015). In den genannten Publikationen beziehen sich die Autoren meist auf eine Einfachstruktur der Dimensionen, d. h. ein Konstrukt mit verschiede-

Tabelle 4.4.: DETECT-Statistik zur Einschätzung der Multidimensionalität eines Tests.

Wert	Interpretation
$DETECT < 0,2$	essentiell unidimensional
$0,2 < DETECT < 0,4$	schwach bis moderat multidimensional
$0,4 < DETECT < 1,0$	moderat bis deutlich multidimensional
$DETECT > 1,0$	stark multidimensional

Anm.: –

nen Fähigkeitsfacetten, von denen jede einzeln in je einer Aufgabe getestet wird. Im Folgenden wird jedoch angenommen, dass der Test unter der Annahme der sogenannten *within-item-dimensionality* (Hartig & Höhler, 2008) analysiert werden sollte (vgl. auch Abschnitt 2.4.4). Dies ist kein Ausschlusskriterium für die Nutzung der DETECT-Statistik (Zhang, 2013), muss aber berücksichtigt werden, weil solch eine Aufgabenstruktur die Werte weniger verlässlich machen kann. Hartig und Höhler (2009) diskutieren die Nutzung multidimensionaler IRT-Modelle in der empirischen Bildungsforschung und betonen die Nützlichkeit im Sinne zusätzlicher Strukturinformationen bei der Testentwicklung. Obwohl viele Schulleistungsstudien „oftmals aus pragmatischen Gründen“ (Mayer & Wellnitz, 2014) mit eindimensionalen Konstrukten arbeiten, ist es gleichzeitig plausibel anzunehmen, dass in einer kompetenzorientierten, naturwissenschaftlichen Aufgabe mehrere Fähigkeitsdimensionen einen Einfluss haben (ebd., vgl. außerdem Abschnitt 2.4.4 und 4.2.2). Bei mehreren, in den Aufgaben verschachtelten Dimensionen, wird die Interpretation von Aufgabenparametern deutlich erschwert. Aus diesem Grund werden für die Hauptstudie auch *multidimensionale Aufgabenschwierigkeiten* (MDIFF), sowie *multidimensionale Aufgabentrennschärfen* (MDISC) berichtet (vgl. Paek & Cole, 2020; Reckase, 2009). Diese erlauben einen übersichtlichen Zugang auf die entsprechenden Parameter, allerdings ohne, dass weitere Fähigkeiten im Rahmen eines Hintergrundmodells berücksichtigt werden können. Das hat den Vorteil, dass kompensatorische Effekte zwischen den eventuellen Dimensionen auftreten können. Gleicht man dann die MDIFFs und MDISCs mit den entsprechenden Werten aus anderen Skalierungen ab, so erhält man eine zusätzliche Perspektive auf die Gültigkeit der statistischen Modellannahmen.

4.4.3. Modellgeltungsprüfung

Für die Geltungsprüfung statistischer Modelle stehen eine Reihe von Verfahren zur Verfügung, die für die jeweilige Forschungsfrage ausgewählt und interpretiert werden müssen (Bühner, 2011). In dieser Arbeit werden die folgenden Verfahren verwendet.

Fit-Indices

Aus der Literatur zu Fit-Statistiken in IRT-Modellen ergeben sich verschiedene Kennwerte mit je besserer oder schlechterer Eignung. Einen allgemeinen Überblick geben Bühner (2011) für die KTT, Chon et al. (2010) für komplexere IRT-Modelle und Bond

4. Methoden

und Fox (2015) speziell für das unidimensionale 1-PL-Modell. Unter der Berücksichtigung der vorliegenden Bedingungen, werden *Infit/Outfit*-Werte für die Beurteilung des Modellfits bei den eindimensionalen Skalierungen aus der quantitativen Vorstudie verwendet (vgl. Abschnitt 5.2.2). Der *Infit* wird im Folgenden auch als *Standardized Weighted Mean Square* (wMNSQ) und der *Outfit* als *Standardized Unweighted Mean Square* (MNSQ) abgekürzt. Die Zahlenwerte werden häufig als Maß für die Abweichung zwischen der durch das Modell erwarteten, sowie der empirisch beobachteten Varianz interpretiert. Präziser ist es aber, die Werte als Antwort auf die Frage nach der *relativen* Aufgabentrennschärfe zu verstehen. Denn obwohl Spannweiten von Fit-Werten als Orientierung für die Aufgabeneignung *unter Gültigkeit* des 1-PL-Modells berichtet werden (Tabelle 4.5), sollten sie besser zum Anlass genommen werden, die *Ursache* für die Nicht-Passung zu suchen, anstatt Aufgaben aufgrund eines einzigen Kennwerts auszuschließen (vgl. Wu et al., 2016, S. 153 ff.). Ein Test, in dem alle Aufgaben einen wMNSQ um 1 haben, ist nicht automatisch als gut zu bewerten. Die angegebene Spannweite drückt lediglich aus, dass *wenn* das 1-PL-Modell gilt, die Aufgaben für eine psychometrische Messung produktiv sind. Insofern sollten diese Statistiken immer im Zusammenspiel mit der Test-/Aufgabenreliabilität und unter Berücksichtigung der Aufgabeninhalte interpretiert werden (Han & Johnson, 2019). Tabelle 4.6 fasst die in der Studie verwendeten

Tabelle 4.5.: Orientierungswerte zur Beurteilung von Infit/Outfit. Verändert nach Bond und Fox (2015).

wMNSQ/MNSQ	Varianzbeurteilung	Interpretation	Misfit-Bezeichnung
> 1,3	zu hoch	weniger prädiktiv	Underfit
< 0,75	zu niedrig	zu deterministisch	Overfit

Anm.: –

Fit-Statistiken zusammen. Neben dem Überblick bei Chon et al. (2010) wurden weitere Quellen herangezogen, um eine möglichst passgenaue Auswahl treffen zu können (Maydeu-Olivares & Joe, 2005, 2006; Orlando & Thissen, 2000, 2003; Stone & Zhang, 2003; Zhang & Stone, 2007). Die Autoren weisen durchgängig darauf hin, dass Fit-Indices immer nur ein Teil der Aufgabenbewertung sein können. Es liegen keine generalisierbaren Strategien vor, bei der die Nutzung solcher Kennwerte automatisch zur Qualitätsverbesserung führt und insofern bietet sich ein pragmatischer Umgang mit ihnen an (Crisan et al., 2017). So werden sie als Orientierungswerte im Zusammenspiel mit dem theoretischen Unterbau sowie dem Ausschluss unplausibler Schlussfolgerungen verwendet.

Item Characteristic Curves (ICCs)

ICCs werden als grafische Hilfsmittel bei der Beurteilung der Güte ein- und zweidimensionaler IRT-Modelle herangezogen. Sie tragen die Lösungswahrscheinlichkeit für eine Aufgabe gegen die Personenfähigkeit ab. Dabei wird dann zunächst die theoretische,

Tabelle 4.6.: Verwendete Fit-Indices.

Ziel der Prüfung	Bezeichnung	Quelle
Homogenität der Trennschärfe in 1-dimensionalen Modellen	<i>Infit/Outfit</i>	Bond und Fox (2015)
relative Modellgeltungsprüfung	$\log\text{Lik}/\chi^2$	Bühner (2011)
absolute Modellgeltungsprüfung	M2	Maydeu-Olivares und Joe (2005)

Anm.: –

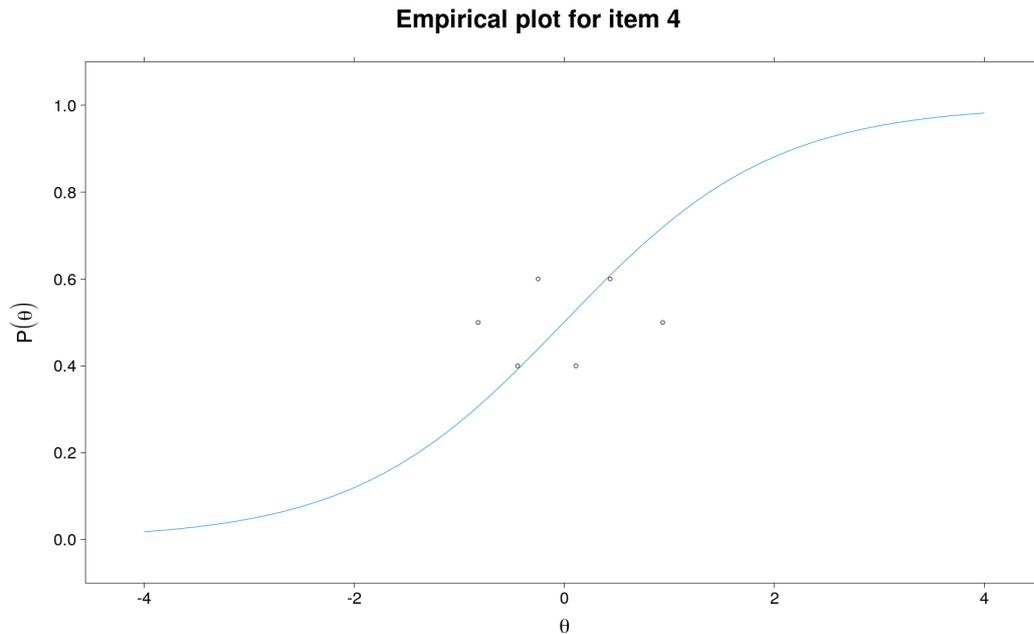


Abbildung 4.4.: ICC-Beispiel mit Misfit.

durch das angenommene Modell vorhergesagte, Kurve abgetragen. Anschließend werden die empirisch vorliegenden Lösungswahrscheinlichkeiten bei geschätzter Personenfähigkeit aus frei wählbaren Fähigkeitsintervallen in das Koordinatensystem eingefügt. So kann die Vorhersagekraft des Modells eingeschätzt werden, indem geprüft wird, ob das Modell und die Daten dicht beieinander liegen. Abbildung 4.4 zeigt eine Kurve mit eindeutig nicht akzeptabler Passung. Das generische Modellkurvenbeispiel enthält die zugehörigen Schätzwerte in sechs Fähigkeitsintervallen. Die Schätzwerte und die Modellkurve zeichnen keinen gemeinsamen Trend und das Item sollte inhaltlich und psychometrisch überprüft werden. Abbildung 4.5 zeigt eine Kurve mit guter Passung. Die Modellkurve und die Schätzwerte folgen einem ähnlichen Verlauf.

Für die Bewertung der ICCs gibt es keine objektiven Maße (Wu et al., 2016). Sie sind ins Verhältnis zu den anderen Modellgütekriterien und inhaltlichen Erwägungen zu setzen. Darüber hinaus können sie nur für Modelle mit höchstens zwei angenommenen Fähigkeitsdimensionen eingesetzt werden. Für höherdimensionale Konstrukte funktio-

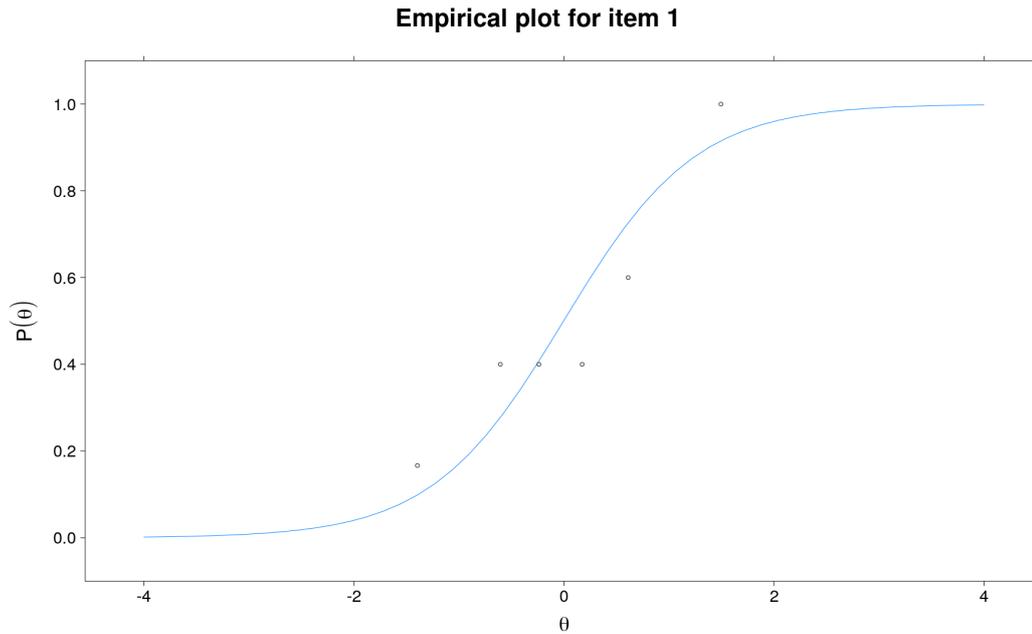


Abbildung 4.5.: ICC-Beispiel mit guter Passung.

nieren sie nicht mehr und die Beurteilung ist auf der Basis von Fit-Statistiken (s. o.) und relativen Modellvergleichen (s. u.) vorzunehmen. Für die vorliegende Studie sind sie dennoch ein nützliches Hilfsmittel, um bspw. die Passung der IRT-Skalierung der kognitiven Fähigkeiten zu prüfen und auch um, unter der Annahme von Unidimensionalität der Modellnutzungsfähigkeit, eine grundsätzliche Prädiktionskraft der Aufgaben abschätzen zu können.

Modellvergleiche

In dieser Studie sind die statistischen Modelle ineinander verschachtelt. Das bedeutet, dass alle beschriebenen mehrdimensionalen Modelle als komplexere Spezialfälle des eindimensionalen Modells aufgefasst werden. Bei der Parameterschätzung mit ML-Methoden wird ein Kennwert, die *logarithmierte Likelihood* (logLik) berechnet. Die *Likelihood Ratio* (LR), also das arithmetische Verhältnis dieser Zahlen, folgt einer Chi-Quadrat- (χ^2) -Verteilung und ist so einem χ^2 -Differenztest zugänglich, um eine nicht-zufällig bessere Passung desjenigen Modells zu prüfen, das eine komplexere Annahme der zugrundeliegenden Struktur macht. So kann bspw. entschieden werden, ob das 2-PL-Modell, das sowohl die Aufgabenschwierigkeit und -trennschärfe schätzt, eine signifikant bessere Passung auf die Daten zeigt als das 1-PL-Modell, das nur die Aufgabenschwierigkeiten schätzt. Können die Modelle auf der Grundlage des Tests nicht voneinander unterschieden werden, ist nach dem Prinzip der Modellsparsamkeit dasjenige Modell als gültig auszuwählen, das die geringeren Zusatzannahmen macht. Solch ein Sparsamkeitsurteil wird auf der Grundlage von *Informationskriterien* gefällt, die ein numerischer Ausdruck für das Verhältnis der logLik zur Anzahl der geschätzten Parameter sind, indem sie

Gewichtungen durch Strafterme für zusätzliche Modellparameter einfügen. Tabelle 4.7 zeigt die hier verwendeten Informationskriterien, zusammengefasst nach Bühner (2011) und erweitert nach Burnham und Anderson (2004). Dabei werden die dargestellten Vergleichsmaße zwar in die Überlegungen mit einbezogen, sie stellen aber nicht das primäre Entscheidungskriterium dar (vgl. auch Ziepprecht et al., 2017), weil sie für Stichproben in der Größenordnung der Studie keine hinreichend akkuraten Aussagen treffen können (Henson et al., 2007). Burnham und Anderson (2004) verweisen auf die Nutzbarkeit von

Tabelle 4.7.: Verwendete Informationskriterien.

Bezeichnung	Berechnung ¹
<i>Akaike Information Criterion</i> (AIC)	$AIC = -2 * LL + 2 * p$
<i>bias corrected AIC</i> (AICc)	$AICc = -2 * LL + 2 * p \frac{p+1}{n-p-1}$
<i>Bayesian Information Criterion</i> (BIC)	$BIC = -2 * LL + \ln(n) * p$
<i>sample size adjusted BIC</i> (saBIC)	$saBIC = -2 * LL + \ln \frac{n-2}{24} * p$

Anm.: ¹LL = LogLik, n = Stichprobengröße, p = Parameteranzahl.

Informationskriterien unter der Beachtung des Verhältnisses der Anzahl geschätzter Parameter zur erhobenen Fallzahl und lassen den Wert des AIC als für diese Studie nur eingeschränkt nutzbar erscheinen. Dieser kann nur dann ohne Einschränkungen genutzt werden, wenn die Anzahl der Fälle pro Parameter $N > 40$. Wenn das nicht zutrifft, empfehlen die Autoren den AICc als korrigierten Kennwert. Darüber hinaus sollte ein Modellvergleich nicht an absoluten Werten dieser Informationskriterien vorgenommen, sondern sog. Modellgewichte (w_i) berechnet werden. Diese nehmen Zahlwerte zwischen 1 und 0 ein und können als Wahrscheinlichkeit für die Auswahl des am besten geeigneten Modells interpretiert werden. In der vorliegenden Arbeit werden Zufallsmodelle geschätzt (vgl. Abschnitt 4.2.1) und über die Modellgewichte des AICc mit dem theoriekonformen Modell verglichen. Gleichung 4.5 zeigt die Berechnung, die in Abschnitt 5.3.3 angewendet wird.

$$w_i = \frac{e^{-\Delta_i * 0.5}}{\sum_{r=1}^R e^{-\Delta_r * 0.5}} \quad (4.5)$$

Dabei gilt

$$\Delta_i = AICc_i - AICc_{min} \quad (4.6)$$

mit $AICc_i$ als den jeweiligen Werten für R verschiedene, interessierende Modelle und $AICc_{min}$ als kleinstem Wert der zu vergleichenden Modelle.

4.5. Zusammenfassung des Methodenteils

Im Hinblick auf die theoretischen Annahmen über die Struktur der Modellnutzung und mit den Testgütekriterien empirisch-didaktischer Forschungsvorhaben als Prüfinstanzen,

4. Methoden

wurden MC-Aufgaben als zentrales Erhebungsinstrument konstruiert und in Testheften zusammengestellt. Der Datenmanagementplan sicherte die Konsistenz, Verfügbarkeit und Sicherheit der erhobenen, personen- und institutionsbezogenen Daten nach der Maßgabe der Datenschutzbestimmungen des Landes Berlin. Die verwendeten statistischen Methoden wurden unter der zugrundeliegenden Fragestellung der Arbeit nach der Dimensionalität des theoretischen Konstrukts ausgewählt. Die Skalierung der Messwerte, d. h. Richtig/Falsch-Antworten, erfolgte primär mit Methoden der IRT, deren Annahmen es erlauben, auch Datensätze mit relativ hohen Missingraten zu bearbeiten. Dies löst nicht automatisch alle Probleme, die Missings mit sich bringen (Bolsinova & Maris, 2016), gewährt aber eine Datenauswertung mit Kennwerten, die ausreichende Rückschlüsse für die Beantwortung der Forschungsfragen zulässt. Die verschiedenen theoretischen Strukturen, werden im Hinblick auf ihre Prädiktionskraft sowie im relativen Vergleich gegeneinander abgewogen. Dazu werden mit Modellgeltungskriterien wie Fit-Indices und Informationskriterien verschiedene Perspektiven zur Beurteilung der jeweiligen Passung erzeugt und in der Gesamtschau für die Strukturaufklärung nutzbar gemacht.

5. Ergebnisse

In diesem Kapitel werden die Resultate zur Bearbeitung der Forschungsfragen dargestellt und es wird in vier Abschnitte gegliedert. Abschnitt 5.1 zeigt die *qualitativen* Vorarbeiten, in denen Expertenratings und offene Schülerfragebögen untersucht wurden und Abschnitt 5.2 fokussiert auf eine *quantitative* Vorstudie zur Qualitätsentwicklung der Aufgaben. Abschnitt 5.3 schließt den empirischen Teil ab, indem es die Skalierungen der verschiedenen statistischen Modelle, sowie deren Vergleiche, zeigt. Bei der Darstellung der Vorarbeiten wird auf die Konsequenzen für die Überarbeitung der Aufgaben hingewiesen. Abschnitt 5.4 schließt den Ergebnisteil mit der datenbezogenen Beantwortung der Forschungsfragen ab.

5.1. Qualitative Vorarbeiten

Die qualitativen Vorarbeiten umfassten ein Rating mit ExpertInnen aus Schulpraxis und Forschung, sowie eine Pilotierung mit offenen Aufgaben bei SuS.

5.1.1. Expertenrating

Es wurden Modelle konstruiert und in ihnen theoriegeleitet Modellkomponenten (vgl. Abschnitt 2.2.2) identifiziert. Die Komponenten wurden benannt (bspw.: ein Elektron) und ExpertInnen ($N = 7$) vorgelegt. Diese sollten, mithilfe eines kurzen Beschreibungstexts der vier Komponenten, Zuordnungen vornehmen, also bspw. entscheiden ob ein Elektron eine kleinste Sinneinheit, eine Relation, eine Operation, oder eine logische Regel darstellt.

Die Interrater-Reliabilität zeigte über alle Modelle und Komponenten eine Spannweite von $\kappa_{min} = .43$ bis $\kappa_{max} = .82$. Niedrige Reliabilitäten wurden – nach Rücksprache mit den Ratern – durch zwei Punkte erzeugt. Erstens schien die Trennung zwischen *Relationen* und *Operationen* in den Beschreibungen nicht eindeutig genug zu sein. Die theoretische Zuordnung wurde aber im Nachgang einstimmig akzeptiert. Der zweite Punkt fiel deutlich stärker ins Gewicht: Es wurde darauf hingewiesen, dass die Modellkomponenten ineinander geschachtelt seien und damit sei bspw. auch die Zuordnung zwischen *kleinsten Sinneinheiten* und *Relationen* erschwert. Eine Relation, so die zusammengefassten Aussagen, könne sinnvollerweise nicht ohne ihre Bezugspunkte benannt werden.

Konsequenzen für Folgeuntersuchungen

Die hierarchische Schachtelung der Komponenten war zwar theoretisch erwartbar, ihr starker Einfluss in einer so frühen Projektphase deutet aber daraufhin, dass dieser Aspekt

besondere Aufmerksamkeit erforderte. Insofern wurde er in Folge des Ratings stets in die Planung aufgenommen. In der Pilotierung sollten aber SuS bewusst Aufgaben ohne ein Prompting auf die Zusammenhänge der Komponenten bearbeiten. Dieser offenere Zugang sollte es ermöglichen, möglichst unbeeinflusste Einsichten in die Herangehensweise der Zielgruppe für die Hauptuntersuchung zu erhalten.

5.1.2. Pilotierung durch offene Aufgaben

Es wurden SuS der 9. Jahrgangsstufe ($N = 42$) offene Aufgaben vorgelegt, in denen Sie Modelle (bspw. Darstellungen von Atomen oder Salzkristallgitter) beschreiben sollten. Sie wurden außerdem gezielt dazu aufgefordert, die Modellkomponenten zu identifizieren. Dazu wurden diese stärker alltagssprachlich umbenannt (Tabelle 5.1). Die 9. Jahrgangsstufe wurde gewählt, um abschätzen zu können, ob die Untersuchung auch mit SuS mit vergleichsweise wenig Lerngelegenheiten in Chemie sinnvoll durchgeführt werden könnte. Die Nennungen wurden im Rahmen einer Masterarbeit codiert, mit den theoretischen

Tabelle 5.1.: Umbenennung der Modellkomponenten in Alltagssprache.

Name (Theorie)	Name (alltagssprachlich)
kl. Sinneinheiten	einzelne Bestandteile
Relationen	Verhältnisse
Operationen	Prozesse
Regeln	Regeln

Anm.: –

a-priori-Zuordnungen verglichen und ausgezählt. Tabelle 5.2 zeigt die Häufigkeitsverteilung. Die SuS benannten mit Abstand meist die kleinsten Sinneinheiten, auch wenn sie

Tabelle 5.2.: Auszählung der Häufigkeiten von den als theoriekonform codierten Zuordnungen durch die SuS.

Modellkomponente	Häufigkeit (%)
kl. Sinneinheiten	78
Relationen	47
Operationen	NA
Regeln	6

Anm.: –

in der Aufgabe explizit zur Suche nach anderen Komponenten aufgefordert wurden. In der offenen Abfrage zu Kommentaren wurde seitens einer Person explizit darauf hingewiesen, dass „[...] für Verhältnisse ja automatisch auch [E]inzelteile [...]“ zu benennen seien, dies sei „[...] irgendwie verwirrend [...]“. Bemerkenswert ist die laut Codiervorlage

nicht vorgenommene Zuordnung von Operationen, d. h. prozessorientierten Nennungen. Weil sie für die Experten (s. o.) aber durchaus identifizierbar waren, wurde dies als Trennschärfeproblem in der Formulierung, sowie als Vertrautheitsproblem interpretiert.

Konsequenzen für Folgeuntersuchungen

Um die curriculare Validität der Hauptstudie zu gewährleisten, wurde nach der Pilotierung entschieden die Grenze für eine Teilnahme der SuS in die 10. Jahrgangsstufe zu verschieben. Dies wurde durch den genannten Mangel an Vertrautheit begründet. Zusätzlich sollte in der Hauptstudie auch ein Modell aus der organischen Chemie verwendet werden. Die Lerngelegenheiten für die OC sind bis einschließlich zur 9. Klasse in Berliner Schulen allerdings nicht systematisch verankert. Bodeneffekte sollten vermieden werden. Analog zum Expertenrating stellte sich die scharfe Trennung der Modellbestandteile in der konkreten Umsetzung auch bei SuS als nicht vollständig möglich dar. Es war in der Folge sowohl aus der theoretischen Perspektive heraus, als auch durch die empirischen Vorprüfungen an zwei verschiedenen Akteursgruppen, eine Multidimensionalität im Sinne einer *Within-Item-Dimensionality* (vgl. Abschnitte 4.4.2 und 2.4.4) zu erwarten.

Es gab keine expliziten Verständnisschwierigkeiten seitens der SuS. Daher wurde die Umbenennung der Modellkomponenten für alle weiteren Untersuchungen beibehalten und davon ausgegangen, dass der Fokus von TestteilnehmerInnen entsprechend gelenkt werden konnte.

5.2. Quantitative Vorarbeiten

Die Vorstudie wurde im Zeitraum Dezember 2016 bis März 2017 an drei Schulen durchgeführt (vgl. Abschnitt 4.1). Die insgesamt 80-minütige Erhebung fand klassenweise ($N = 11$), in 9 von 11 Fällen in Gegenwart einer Lehrkraft statt. Insgesamt nahmen $N = 269$ SuS an der Erhebung teil.

5.2.1. Deskriptive Darstellung der Rohdaten

Tabelle 5.3 stellt die aggregierten Rohpunktmittelwerte mit Standardabweichungen der StudienteilnehmerInnen getrennt nach den Schulen und nach der Entscheidung für Fallauschlüsse (s. u.) dar. Daraus resultierte ein Datensatz mit $N = 249$ TeilnehmerInnen. Es sind die prozentualen Missingquoten hinzugefügt, die als MCAR beurteilt wurden (vgl. Abschnitt 4.3.2). Die fehlenden MCAR in den Modellaufgaben sind dadurch zu erklären, dass keine individuelle Aufgabenbearbeitung vorgesehen war, sondern pro Aufgabe immer eine feste Zeit (2 Minuten) mit anschließendem, gemeinsamen Umblättern vorgegeben wurde.

Missings und Datenausschlüsse

In einer Klasse, in der die Lehrkraft sich nach Beginn der Erhebung aus dem Raum entfernte, musste der Test abgebrochen werden. Die SuS boykottierten den Test und es

Tabelle 5.3.: Anzahl der TeilnehmerInnen sowie Datenmittelwerte der Vorstudie, aufgeschlüsselt nach Schulen, inkl. Datenausschlüssen & MCAR-Quoten.

Variable	ISS ¹	Gymnasium I	Gymnasium II	Missing-%
$N(TN)^{2,3}$	J: 51 M: 47 NA: 2	J: 26 M: 27 NA: 2	J: 46 M: 48 NA: –	– – –
∅ Alter	15.6(.8) ⁴	15.4(.7)	15.2(.4)	1.6
∅ Chemienote	3.3(1.2)	2.7(.9)	3.2(1.1)	2.8
∅ Deutschnote	2.9(1.0)	3.1(.8)	2.5(1.0)	2.8
∅ Mathenote	3.5(1.0)	3.1(.9)	2.5(.9)	2.4
∅ KFT ⁵	9.9(5.6)	15.6(5.0)	17.8(4.0)	< .1
∅ Leseverstehen	5.6(4.4)	10.7(5.5)	14.9(5.7)	2.4
∅ Lesegeschw.	583.0(248.7)	666.9(196.7)	779.4(190.9)	7.6
∅ Fachwissen ⁶	2.7(1.7)	4.5(1.4)	4.8(1.1)	–
∅ Modellnutzung ⁷	3.4(1.5)	4.3(2.1)	6.1(2.0)	–

Anm.: ¹Integrierte Sekundarschule. ²Anzahl TeilnehmerInnen. ³J = Jungen, M = Mädchen, NA = keine Angabe. ⁴Standardabweichung in Klammern. ⁵Punkte_{max}: 25. ⁶Punkte_{max}: 8.

⁷Punkte_{max}: 12. Alle Mittelwerte und ihre Standardabweichungen wurden auf eine Nachkommastelle gerundet.

musste davon ausgegangen werden, dass deswegen keine Vergleichbarkeit mit den anderen Gruppen hergestellt werden konnte. Darüber hinaus gab es vereinzelt qualitative Beobachtungen während der Testungen (bspw. das gelangweilte Einreißen einzelner Testseiten), oder bei der Dateneingabe im Nachgang (bspw. obszöne Zeichnungen oder das durchgehende Ankreuzen einer bestimmten Multiple-Choice-Position), die zur Markierung der Testhefte und nach anschließender Prüfungs zum Fallausschluss führten. Zwei der Fachwissensaufgaben mussten aus dem Datensatz entfernt werden. Sie enthielten Distraktoren, die nicht vollständig disjunkt waren und daher keine eindeutige Lösung erlaubten. In den Zusammenhangsuntersuchungen zwischen den Fähigkeitschätzungen aus den Modellnutzungsaufgaben und den Kovariablen mit Plausible Values (vgl. Abschnitte 4.4.1 und 5.2.2) wurden desweiteren 7 Datensätze per vollständigem Ausschluss entfernt. Die genannte Analyse setzt vollständige Datensätze der Kovariablen voraus. Da das Ziel lediglich eine Trendabschätzung ohne Anspruch auf hypothesengeleitete, statistische Inferenz war, wurde von Imputationsverfahren abgesehen.

5.2.2. Statistische Skalierungen des Fachwissens- und des Modelltests

Dieser Teilabschnitt dient der Modellgeltungsprüfung des Fachwissenstests, sowie der Aufgaben- und Personenparameterschätzung der Modellaufgaben. Dies trägt zur Absicherung der internen Validität der Konstrukte bei (vgl. Abschnitt 4.2.1). Außerdem wurde bei den Modellaufgaben eine explorative Dimensionalitätsprüfung per DETECT-Statistik vorgenommen (vgl. Abschnitt 4.4.2), sowie eine explorative Faktorenanalyse unter dem IRT-Paradigma für die Fachwissensaufgaben angewendet.

Fachwissenstest

Die binär richtig/falsch codierten Aufgaben wurden als unidimensional angenommen und es wurde ein 1-PL-, sowie ein 2-PL-Modell auf die Daten angewendet. Weil ein vollständiger Datensatz vorlag, konnten ein absoluter Modellfit für jedes einzelne Modell, sowie Modellvergleiche eingesetzt werden, um die Gültigkeit zu prüfen (vgl. Abschnitt 4.4.1). Tabelle 5.4 zeigt die absoluten Anpassungsmaße für beide Modelle. Im

Tabelle 5.4.: Vergleich des absoluten Modellfit für 1-PL- & 2-PL-Modell des Fachwissenstests der Vorstudie.

Modell	M2	DF	p-Wert	RMSEA	TLI	CFI
1-PL	45	27	< .05	> .05	< .95	< .95
2-PL	13	20	> .05	< .05	> .95	> .95

Anm.: –

Fall des absoluten Fit, wird das 1-PL-Modell statistisch signifikant zurückgewiesen und das 2-PL-Modell kann beibehalten werden. Tabelle 5.5 zeigt den Modellvergleich. Dabei

5. Ergebnisse

passt das 2-PL-Modell bei größerer Modellsparsamkeit signifikant besser auf die Daten. Es ist aus theoretischen Überlegungen heraus sinnvoll anzunehmen, dass der Fachwis-

Tabelle 5.5.: Relativer Modellvergleich zwischen 1-PL- & 2-PL-Modell des Fachwissens-tests der Vorstudie.

Modell	LogLik (df)	AIC	BIC	LR-Test
1-PL	-1239.53 (240)	2497	2529	–
2-PL	-1223.87 (233)	2480	2536	$\Delta_{\chi^2}(7) = 31.3^{***}$

Anm.: *** $p < .001$.

senstest unterschiedliche Fähigkeitsdimensionen erfordert (Woitkowski, 2015), bzw. dass das deklarative Fachwissen in unterschiedliche Kategorien fällt. Es könnte sein, dass Fragen zu submikroskopischen Strukturen (z. B. Eigenschaften von Elektronen) und Fragen zu Phänomenen (z. B. Aggregatzustandsänderung) in je verschiedene Subdimensionen einzuordnen sind. Aus diesem Grund wurde eine explorative Faktorenanalyse unter dem IRT-Ansatz (Chalmers, 2012) mit 2 Faktoren spezifiziert und mit der einfaktoriellen Skalierung verglichen (Tabelle 5.6). Das zweifaktorielle Modell wird bei den gegebenen

Tabelle 5.6.: Relativer Modellvergleich zwischen der einfaktoriellen und der explorativen, zweifaktoriellen 2-PL-Skalierung des Fachwissenstests der Vorstudie.

Faktor	LogLik (df)	AIC	BIC	LR-Test
1	-1223.87 (233)	2480	2536	–
2	-1220.97 (226)	2488	2569	$\Delta_{\chi^2}(7) = 5.7$

Anm.: –

Daten nicht signifikant, weswegen nach dem Prinzip der Modellsparsamkeit das einfaktorielle Modell beibehalten werden sollte. Aus den oben dargestellten Ergebnissen wurde gefolgert, dass die Fachwissensaufgaben als eindimensionaler Schätzer unter einer 2-PL-Skalierung die beste Erklärung für die Daten liefern. Tabelle 5.7 zeigt schließlich, dass die Schätzung angemessen reliabel ist (Hartmann et al., 2015; Neumann, 2011) und einen ausreichend breiten Fähigkeitsbereich abdeckt.

Gruppenvergleiche über Kovariablen

Zur Prüfung der Gruppenvergleichbarkeit wurden nur die standardisierten Instrumente LGVT, KFT und Fachwissen herangezogen. Die erhobenen Noten fanden keine Berücksichtigung. Aus den Gesprächen mit den jeweiligen Lehrkräften waren viele unterschiedliche fachliche und pädagogische Bewertungsschwerpunkte erkennbar. Aus diesem Grund wurden die Noten als zu stark gruppenspezifisch beurteilt. Weil außerdem die genannten drei Kovariablen in Schulnoten einfließen, diese gleichzeitig aber auch ein

Tabelle 5.7.: Personenfähigkeitsschätzungen für die eindimensionale 2-PL-Skalierung des Fachwissenstests der Vorstudie.

Reliabilität	M (SD)
.63 (WLE)	-0.04 (1.34)
.63 (EAP)	0.00 (.79)

Anm.: –

pädagogisches Urteil beinhalten, wurden sie als weniger informativ bei der Skalierung einer möglichst objektiven Leistungsmessung bewertet. Lesegeschwindigkeit wird durch Leseverständnis moderiert (vgl. O'Reilly & McNamara, 2007; Reed et al., 2017) und das Alter war stichprobenbedingt zu homogen, um bedeutsam zur Varianzaufklärung beizutragen. Gruppenvergleiche zwischen den Schulen durch die Modellaufgaben wurden als nicht valide beurteilt, da diese überhaupt erst in ihrer Struktur untersucht werden sollten. Daher ergab sich die genannte Auswahl und die wie folgt vorgenommenen Transformationen, bzw. Skalierungen der entsprechenden Rohpunktwerte. Die LGVT-Rohwerte wurden zu diesem Zweck in z -Werte transformiert und mit der Normwerttabelle des Instruments (Schneider et al., 2007) abgeglichen. Die KFT-Rohwerte wurden Rasch-skaliert (vgl. Baumert et al., 2009) und für den Fachwissenstest wurden die Personenfähigkeitsschätzer aus der 2-PL-Skalierung verwendet. Nach der, exemplarisch am KFT-Personenfähigkeitsschätzer gezeigten grafischen (Abbildung 5.1) und statistischen Einschätzung über Normalverteilung und Varianzhomogenität (Bortz & Schuster, 2010) wurden die Gruppenvergleiche konservativ mit nicht-parametrischen Methoden vorgenommen (Field et al., 2012). Es ergaben sich eine signifikante Abweichung von der Normalverteilung nach Shapiro-Wilk mit $W = .98, p < .001$ und Varianzhomogenität nach Levene mit $F(2, 244) = 2.79, p = .063$. Ein Kruskal-Wallis-Test zeigte sowohl für den KFT ($H(2) = 70.90, p < .001, \epsilon^2 = .29$), als auch für das Fachwissen ($H(2) = 89.13, p < .001, \epsilon^2 = .37$) und den LGVT ($H(2) = 87.18, p < .001, \epsilon^2 = .36$) signifikante Unterschiede mit großen Effektstärken. Die Post-hoc-Analyse nach Dunn (Dinno, 2015) ergab für kognitive Fähigkeiten den niedrigsten Wert für die integrierte Sekundarschule ($Mdn = -.89$), ein mittleres Fähigkeitsniveau für eines der Gymnasien ($Mdn = .44$) und das höchste Fähigkeitsniveau für das zweite Gymnasium ($Mdn = .84$) bei Signifikanz aller Paarvergleiche ($p < .05$) nach Benjamini-Hochberg-Korrektur (vgl. auch Abbildung 5.2). Die Tests, Gruppenvergleiche und Grafiken für alle drei Kovariablen sind in Anhang C hinterlegt. Die Ergebnisse zeigen eine durchgehend niedrigere Leistungsfähigkeit in der integrierten Sekundarschule. Dies wurde in die weiteren Überlegungen für die Konzeption der Hauptstudie aufgenommen und muss als Hinweis darauf gedeutet werden, die Ergebnisse aus den Aufgaben nur vorsichtig interpretieren zu können.

Verteilung der KFT-Fähigkeitsschätzer über die Schulen

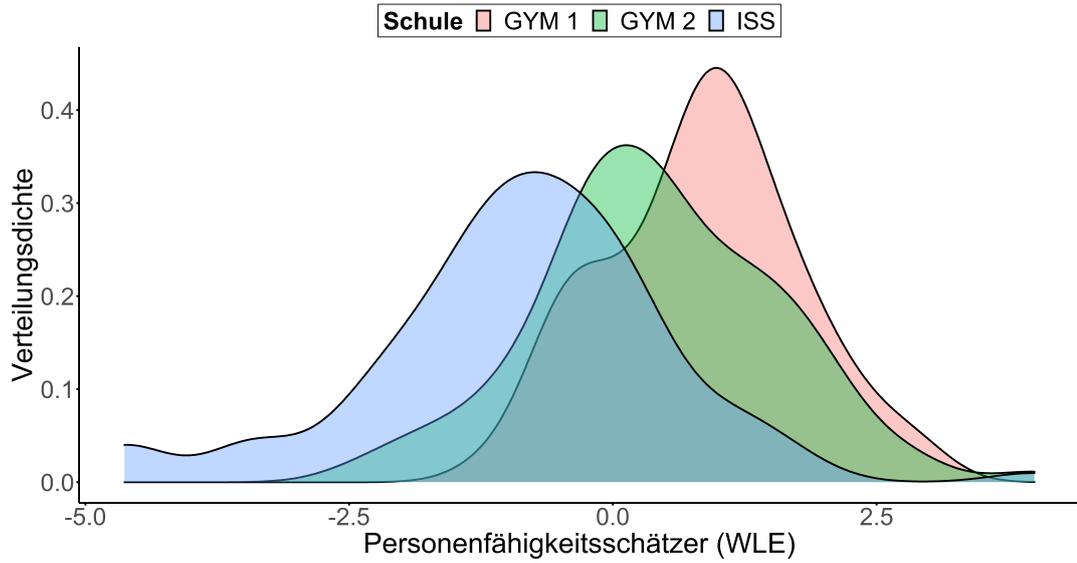


Abbildung 5.1.: Grafische Verteilungsprüfung der Fähigkeitsschätzer im KFT der Vorstudie.

Vergleich der KFT-Fähigkeitsschätzungen zwischen den Schulen

$\chi^2_{\text{Kruskal-Wallis}}(2) = 70.90, p = < 0.001, \hat{\epsilon}^2 = 0.29, \text{CI}_{95\%} [0.20, 0.38], n_{\text{obs}} = 242$

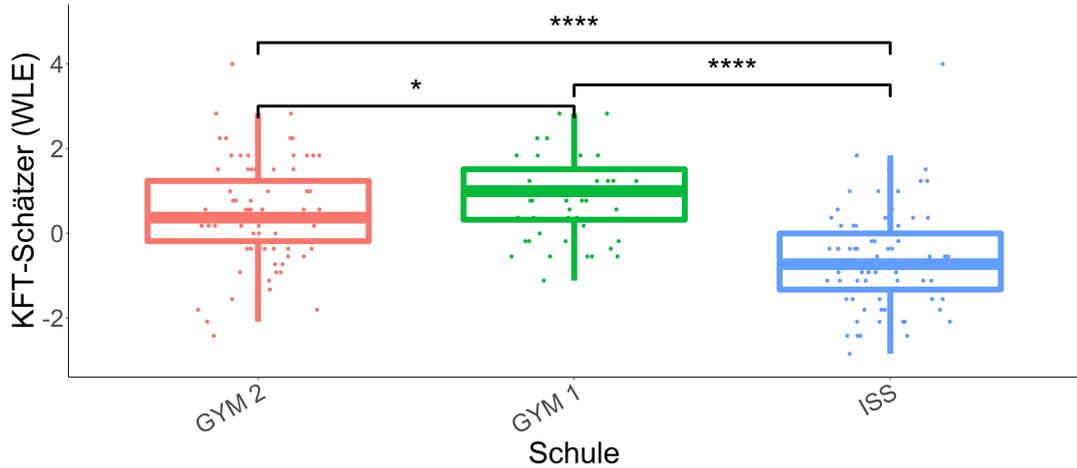


Abbildung 5.2.: Schulvergleich der KFT-Fähigkeitsschätzer in der Vorstudie.

Modellaufgaben

Für die Einschätzung der konstruktbezogenen Aufgaben waren nur relative Modellvergleiche möglich, weil aufgrund des Studiendesigns eine große Anzahl Missings-by-Design vorlagen. Es wurde eine 1-PL- sowie eine 2-PL-Skalierung vorgenommen und beide miteinander verglichen (Tabelle 5.8). Weil das 2-PL-Modell die Daten auf Kosten höherer

Tabelle 5.8.: Relativer Modellvergleich zwischen 1-PL- & 2-PL-Modell der Modellaufgaben in der Vorstudie.

Modell	LogLik (df)	AIC	BIC	LR-Test
1-PL	-1770.71 (188)	3663	3878	–
2-PL	-1692.49 (129)	3625	4047	$\Delta\chi^2(59) = 156.4^{***}$

Anm.: $***p < .001$.

Komplexität besser erklärt, dies aber nur relativ zum 1-PL-Modell aussagekräftig ist, wurden zusätzliche Prüfungen vorgenommen, um die Modellgültigkeit abzusichern. Tabelle 5.9 zeigt die Spannweite der Infit-Werte der Aufgaben unter 2-PL-Skalierung. Die Infit-Werte können, unter Gültigkeit des Modells (vgl. Abschnitt 4.4.1), als sehr gut passend bewertet werden (Bond & Fox, 2015). Der Outfit spiegelt Extremwerte in der Lösungswahrscheinlichkeit wider und schwankt stärker. Dies wird in der einschlägigen Literatur allerdings als unproblematisch betrachtet (vgl. Wu et al., 2016) und darüber hinaus liegen die Werte innerhalb der als akzeptabel beschriebenen Spanne (Bond & Fox, 2015). Aus der Verteilung der EAP-Fähigkeitsschätzungen (Tabelle 5.10) wurden

Tabelle 5.9.: Überblick der Infit/Outfit-Werte für die 2-PL-Skalierung der Modellaufgaben in der Vorstudie.

M (SD)	Infit		M (SD)	Outfit	
	Min	Max		Min	Max
.99(< .01)	.98	1.01	1.00(.06)	.74	1.27

Anm.: –

pro Person 10 Plausible Values gezogen und mit den Fähigkeitsschätzern der Kovariablen korreliert (vgl. Abschnitt 4.4.1). Abbildung 5.3 zeigt exemplarisch den nach den Schulen getrennten Zusammenhang von Modellnutzungsfähigkeit und KFT, unter der Annahme eines einfachen linearen Modells. Die Zusammenhänge mit den anderen Kovariablen, sowie die Produkt-Moment-Korrelationen nach Pearson (Field et al., 2012) sind in Anhang D hinterlegt. Die modellbasierte, multivariate Regression der Kovariablen auf die angenommene Modellnutzungsfähigkeit nach der Methode von Rubin (1987) (Tabelle 5.11) stützt die Annahme der positiven Zusammenhänge aus dem Theorieteil (vgl. Abschnitt 4.2.3). Weil alle Korrelationen statistisch bedeutsam sind, aber ca. 60 % der

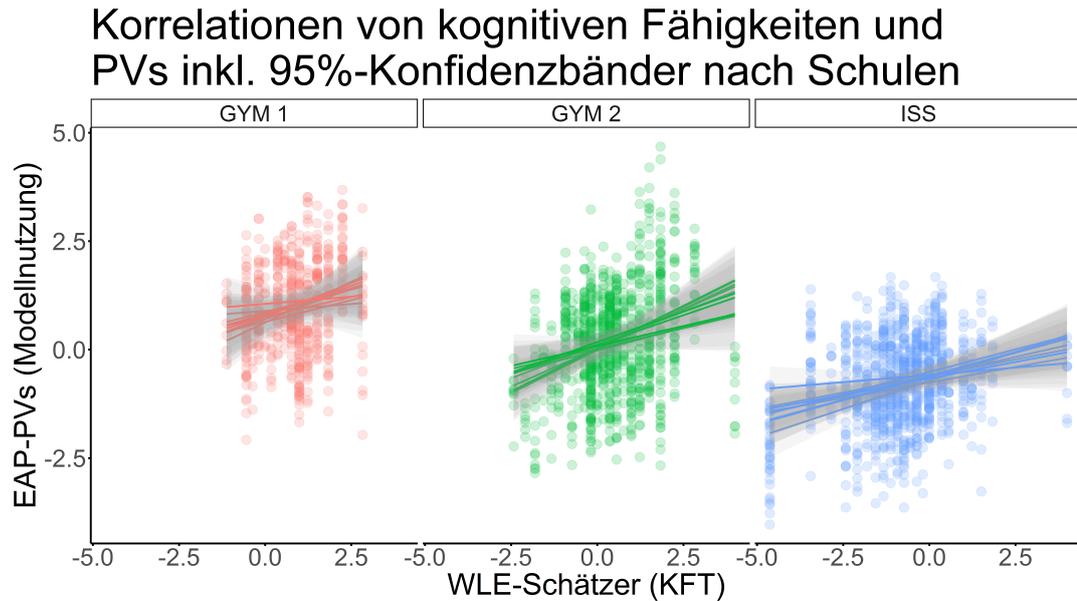


Abbildung 5.3.: Regressionslinien für EAP-PVs der Modellaufgaben über den Fähigkeitsschätzern kognitiver Fähigkeiten getrennt nach Schulen.

Gesamtvarianz unerklärt bleibt, wird die Annahme einer latenten Fähigkeit zur Modellnutzung nicht zurückgewiesen. Für eine weitere Untersuchung sind die ICCs für jede

Tabelle 5.10.: Personenfähigkeitsschätzungen für die eindimensionale 2-PL-Skalierung der Modellaufgaben der Vorstudie.

Reliabilität	M (SD)
.63 (WLE)	0.05 (1.61)
.71 (EAP)	0.00 (.85)

Anm.: –

einzelne Aufgabe überprüft worden. Diese kontrastierten die Annahme, dass das eindimensionale 2-PL-Modell auf die Daten passt: Abbildung 5.4 zeigt exemplarisch eine Aufgabe, bei der die empirischen Fähigkeitsschätzer gut mit der Modellkurve übereinstimmen. Abbildung 5.5 zeigt eine Kurve bei der das ebenfalls der Fall ist – allerdings müsste sie derart interpretiert werden, dass eine höhere Personenfähigkeit zu einer geringeren Lösungswahrscheinlichkeit führt. Das wäre ein schweres Validitätsproblem und musste in die weiteren Überlegungen einbezogen werden. Nach der systematischen Einzelprüfung der Aufgabenstämme und der Distraktoren im Rahmen einer Masterarbeit, wurden Formulierungsschwächen deutlich, die einen Einfluss auf die Lösungswahrscheinlichkeit gehabt haben könnten und die für die Hauptstudie überarbeitet werden mussten. Über diesen formalen Ansatz hinaus, weisen Paek und Cole (2020) sowie Wu et al. (2016) darauf hin, dass systematische Unterschiede in den Steigungen von eindimensionalen 2-

Tabelle 5.11.: Lineare Regression der über die Plausible Values gemittelten Zusammenhänge der 1-dimensional skalierten Modellnutzungsfähigkeit und Kovariablen über alle Schulen.

Modellpassung	R	R^2
	.64	.41
Statistik	β	SE_{β}
Konstante	.006	.071
kogn. Fähigkeiten	.144*	.066
Fachwissen	.267**	.061
Leseverständnis	.467**	.101

Anm.: * $p < .01$, ** $p < .001$.

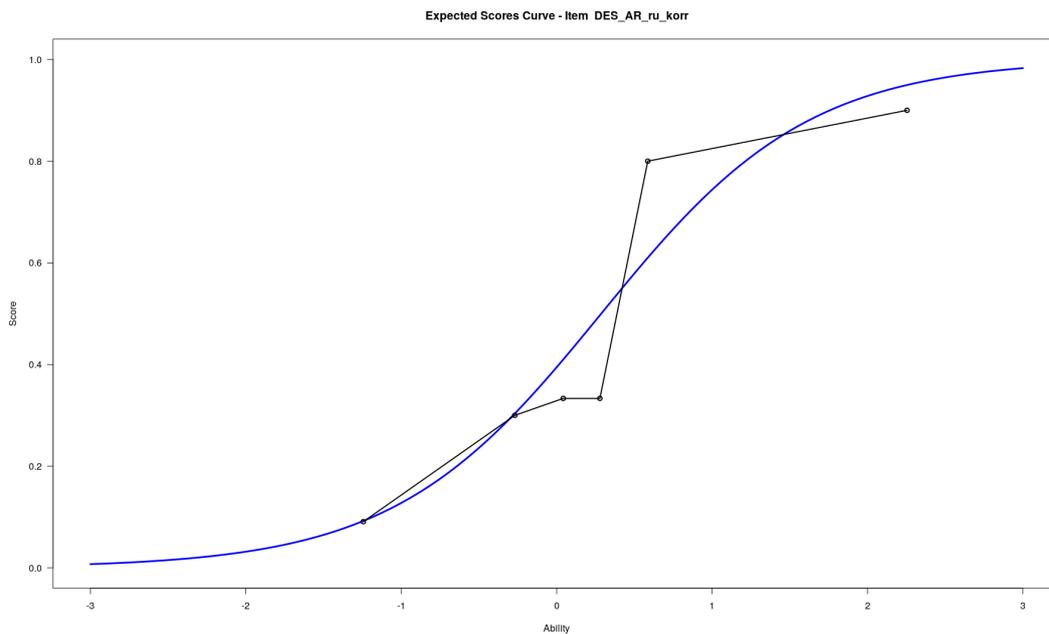


Abbildung 5.4.: ICC der Aufgabe zu *Destillation* mit Denkschritt *Auswertung / Reflexion* und Modellkomponente *Regeln*.

5. Ergebnisse

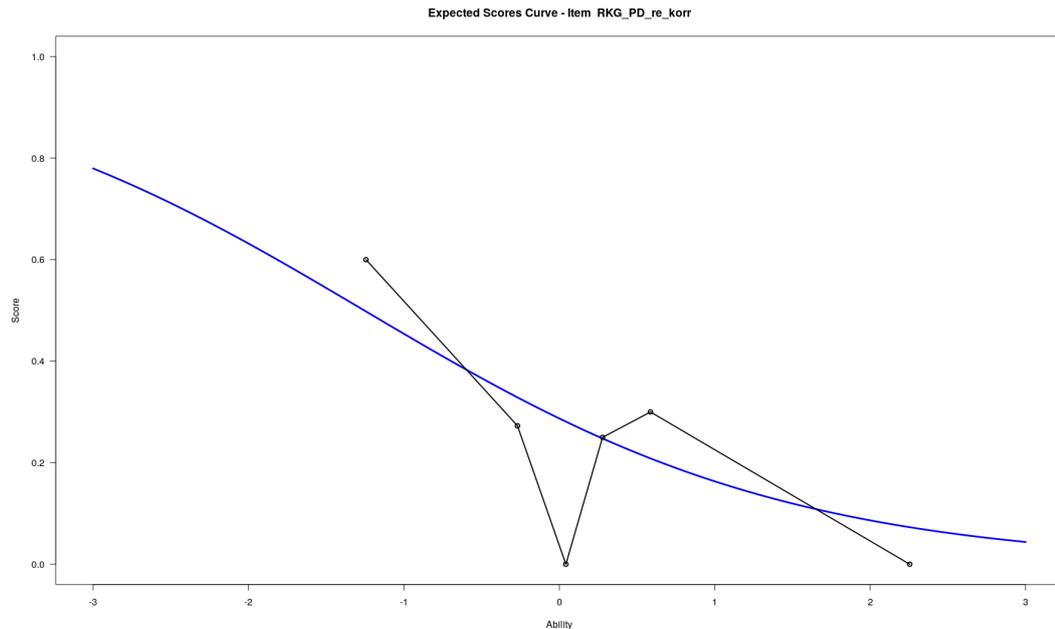


Abbildung 5.5.: ICC der Aufgabe zu *Reaktionsgleichungen* mit Denkschritt *Planung / Durchführung* und Modellkomponente *Relationen*.

PL-Skalierungen (bis hin zu negativen Steigungen, vgl. Chalmers, 2015) in IRT-Modellen auch durch Multidimensionalität zustande kommen können. Im Zusammenspiel mit den qualitativen Vorarbeiten (Abschnitt 5.1), die auf eine Verschachtelung der Modellkomponenten in den Aufgaben hinwiesen, wurde eine explorative Analyse unter Nutzung der DETECT-Statistik (vgl. Abschnitt 4.4.2) vorgenommen. Dazu musste vorab eine Vorgabe über die maximale Zahl der Dimensionen gemacht werden, die theoriegeleitet und begründet sein sollte. Dazu wird auf die Strukturvorschläge aus dem Theorieteil in Abschnitt 2.4.4 verwiesen. Unter der konservativen Vorgabe von vier orthogonalen Dimensionen – d. h. vier Modellkomponenten, mit denen in vier streng voneinander trennbaren Fähigkeitsdimensionen umgegangen wird – wird eine Lösung mit drei Dimensionen ($DETECT_{max} = .74$) bevorzugt. Unter der Annahme von sieben Dimensionen werden sieben Aufgabencluster mit $DETECT_{max} = .88$ zurückgegeben. Dies kann als zugrundeliegende Multidimensionalität gedeutet werden. Inferenz auf konkrete Zuordnungen der Aufgaben zu Dimensionen wurde in diesem explorativen Ansatz nicht vorgenommen. Aufgrund der relativ geringen Fallzahl und des Überarbeitungsbedarfs der Aufgaben (vgl. Abschnitt 4.1), wurden die berechneten Kennwerte nur als Hinweise und nicht als absolute Zuordnungskriterien interpretiert.

5.2.3. Zusammenfassung und Konsequenzen für die Hauptstudie

Die Skalierungen der Aufgaben über alle Schulen zeigten, dass die Übersetzung der theoretischen Überlegungen in Aufgaben grundsätzlich zu einer Messbarmachung führten. Der Fachwissenstest war messproduktiv und auch die Modellaufgaben konnten reliabel

für Messungen genutzt werden. Die starke Abhängigkeit der Leistungen von der Schulform ließ Reliabilitäts- und Validitätseinschränkungen zum Vorschein kommen. Dieser Effekt ist außer in den spezifischen Leistungsdaten der vorliegenden Studie, auch aus Large-Scale-Studien bekannt (vgl. Carstensen et al., 2007). Aus den signifikanten Leistungszusammenhängen und der antizipierten Komplexität des zu prüfenden Konstrukts der Modellnutzung, ergab sich die Entscheidung für die Integration der kognitiven Fähigkeiten, des Fachwissens, sowie des Leseverständnisses in die Modelle der Hauptstudie. Mithilfe einer multiplen, latenten Regression sollte für den Einfluss dieser Variablen kontrolliert werden.

Weil konfirmatorische IRT-Modelle mit verschiedenen Dimensionen skaliert werden sollten und dazu für die einzelnen Aufgaben die Schätzung der Steigungsparameter notwendig war, bedurfte es einer größeren Anzahl TeilnehmerInnen. Weil die Bearbeitungszeit der Vorstudienaufgaben als zu großzügig beurteilt wurde, sollte die Anzahl der Bearbeitungen durch die Erhöhung der Modellaufgaben von 12 auf 15 pro Testheft gesteigert werden. Alle Aufgaben wurden zudem überarbeitet und angepasst, inhaltlich aber nicht neu konstruiert. Für die Hauptstudie wurden außerdem die Fachwissensaufgaben teils neu konstruiert und teils überarbeitet, um eine möglichst gute Passung zu erreichen.

5.3. Die Hauptstudie

5.3.1. Deskriptive Ergebnisse

Insgesamt nahmen $N = 524$ SuS an der Hauptstudie teil. Tabelle 5.12 stellt die aggregierten Rohpunktwerte der StudienteilnehmerInnen nach der Entscheidung für Fallausschlüsse (s. u.) dar. Daraus resultierte ein Datensatz mit $N = 513$ TeilnehmerInnen. Es sind die prozentualen Missingquoten hinzugefügt, die als MCAR, bzw. im Fall der Modellaufgaben als MAR beurteilt wurden (vgl. Abschnitt 4.3.2). Die Erhebung der Hauptstudien Daten fand im Zeitraum Juni bis Oktober 2017 in acht Berliner Gymnasien statt.

Missings und Datenausschlüsse

Die Testung einer Teilstichprobe wurde in Rücksprache mit der Lehrkraft abgebrochen, um die spürbar ermüdeten SuS zu entlasten. Eine Fachwissensaufgabe musste bereits vor der Analyse aus dem Test entfernt werden. Die Distraktoren waren missverständlich formuliert und die Validität der Aufgabe konnte nicht gewährleistet werden.

5.3.2. Skalierung des Fachwissenstests

Die Stichprobe stammte nur aus Gymnasien und die Aufgaben waren infolge der Vorstudie überarbeitet worden. Daher wurde – unter der Annahme einer homogenen Fähigkeitsverteilung als in der Vorstudie – der Fachwissenstest sowohl unter Annahme des 1-PL-Modells, als auch unter der Annahme des 2-PL-Modells skaliert und ein relativer Modellvergleich durchgeführt (Tabelle 5.13). Das 2-PL-Modell wird bevorzugt und zeigt

5. Ergebnisse

Tabelle 5.12.: Erhebungsrohdaten der Hauptstudie nach Bereinigung, inkl. Datenauschlüssen & MCAR/MAR-Quoten.

Variable	M (SD) ¹	SE ²	Mdn	Min	Max	Skew	Kurtosis	Miss. (%) ³
Alter	15.30 (0.73)	0.03	15	14	18	0.66	1.19	.9
Chemienote	2.51 (0.94)	0.04	3	1	6	0.23	-0.16	2.1
Deutschnote	2.56 (0.92)	0.04	3	1	6	0.19	-0.25	1.7
Mathenote	2.61 (0.97)	0.04	3	1	6	0.26	-0.40	1.5
KFT	18.57 (4.75)	0.21	20	0	25	-0.95	0.41	–
Lese- verstehen	13.81 (5.81)	0.26	14	0	33	0.28	0.28	–
Lese- geschw.	738.62 (206.63)	9.27	720	0	1727	0.75	1.47	3.1
Fachwissen ⁴	5.69 (1.91)	0.08	6	0	9	-0.17	-0.70	–
Modell- aufgaben ⁵	5.58 (2.67)	0.12	5	0	14	0.33	-0.24	–

Ann.: ¹Arithmetisches Mittel, Standardabweichung in Klammern. ²Standardfehler. ³Werte auf eine Nachkommastelle gerundet. ⁴Erreichbare Punkte_{max}: 9. ⁵Erreichbare Punkte_{max}: 15.

Tabelle 5.13.: Relativer Modellvergleich zwischen 1-PL- & 2-PL-Modell des Fachwissens-tests der Hauptstudie.

Modell	LogLik (df)	AIC	BIC	LR-Test
1-PL	-2708.73 (503)	5437	5480	–
2-PL	-2692.77 (495)	5421	5498	$\Delta_{\chi^2}(8) = 31.9^{***}$

Ann.: $***p < .001$.

bei größerer Modellsparsamkeit eine akzeptable Reliabilität (Tabelle 5.14). Die Reliabilität wird als akzeptabel beurteilt, weil curriculare Unterschiede zwischen den einzelnen Gymnasien eine homogenere Fähigkeitsverteilungsannahme auch innerhalb der gleichen Schulform konterkarieren. Diese Schuleffekte werden in Large-Scale-Studien üblicherweise durch die hohe Probandenanzahl kompensiert (Wu et al., 2016, S. 41 ff.), können aber in kleineren Studien zum beobachteten Wert führen, obwohl die Probandenzahl relativ zur geschätzten Parameterzahl hoch ist (Wu, 2010). Der Infit/Outfit spricht unter der Annahme des eindimensionalen 2-PL-Modells dafür, dass der Test produktiv im Sinne einer Fähigkeitsmessung ist (Tabelle 5.15, vgl. Abschnitt 4.4.3). Die erhaltenen WLE-Schätzer wurden für das geplante Hintergrundmodell der komplexeren Dimensionalitätsprüfungen der Modellaufgaben eingesetzt (vgl. Abschnitt 5.3.3).

Tabelle 5.14.: Personenfähigkeitsschätzung (WLE) für die eindimensionale 2-PL-Skalierung des Fachwissenstests der Hauptstudie.

Reliabilität	M (SD)
.64	0.03 (1.35)

Anm.: –

Tabelle 5.15.: Überblick der Infit/Outfit-Werte für die eindimensionale 2-PL-Skalierung des Fachwissenstests der Hauptstudie.

M (SD)	Infit		Outfit		
	Min	Max	M (SD)	Min	Max
1.00(< .01)	.99	1.01	.99(.02)	.93	1.01

Anm.: –

5.3.3. Skalierungen der Modellaufgaben

Reckase (2009) empfiehlt bei der Untersuchung von kognitiven Konstrukten ein sukzessives Vorgehen von einfachen zu komplexeren Modellen. Wie in der Vorstudie gezeigt werden konnte, liegt grundsätzlich eine Messbarkeit der Fähigkeit zur Lösung der Aufgaben vor (Abschnitt 5.2). Mit den Daten aus der neuen Stichprobe wurden die Aufgaben erneut unter verschiedenen Strukturannahmen skaliert und die folgenden Unterabschnitte beschreiben die erhaltenen Kennwerte mit jeweils ansteigender Komplexität. Im Anschluss werden Modellvergleiche dargestellt, um dasjenige Modell zu identifizieren, das am besten auf die Daten passt.

Eindimensionales Modell

Die Zusammensetzung der Stichprobe wurde im Übergang von den Vorarbeiten hin zur Hauptstudie verändert, sodass die Aufgaben auf grundsätzliche Messproduktivität überprüft werden mussten. Unter der Annahme von einer einzigen Fähigkeitsdimension, wurden darum sowohl ein 1-PL-, als auch ein 2-PL-Modell spezifiziert. Die designbedingten Missings ließen absolute Modellfitwerte (vgl. Abschnitt 5.2) nicht zu, sodass auf Basis der ICCs, der Infit-/Outfitstatistik (Tabelle 5.16) und relativer Modellvergleiche eine Auswahl der besten Passung gewählt wurde. Darüber hinaus wurden als Fähigkeitsschätzer EAP-PVs (vgl. Abschnitt 4.4.1 sowie Wu et al., 2016) verwendet. Die in der Vorstudie geschätzten Aufgabenschwierigkeiten wurden als Initialwerte in die Schätzung für die Hauptstudie aufgenommen, um die Berechnung effizienter zu machen. Tabelle 5.17 zeigt den Modellvergleich. Auf Kosten höherer Komplexität erklärt das

Tabelle 5.16.: Überblick der Infit/Outfit-Werte für die eindimensionale 2-PL-Skalierung der Modellaufgaben in der Hauptstudie.

M (SD)	Infit		Outfit		
	Min	Max	M (SD)	Min	Max
1.00 (< .001)	.99	1.01	1.00 (.02)	.97	1.05

Anm.: Alle Werte wurden auf die zweite Nachkommastelle gerundet.

2-PL-Modell die Daten besser. Wie in der quantitativen Voruntersuchung treten teils negativen Steigungsparameter bei der 2-PL-Schätzung auf (vgl. Abschnitt 5.2.2). Damit wird seine Nutzung als restriktives Basismodell für die folgenden, höherdimensionalen Fälle insgesamt als gerechtfertigt angenommen, weil eine Dimensionalitätszuordnung aufgrund eben dieser Steigungen erfolgt (Abschnitt 4.4.1). Weil die Nutzung von Hinter-

Tabelle 5.17.: Relativer Modellvergleich zwischen 1-PL- & 2-PL-Modell der Modellaufgaben.

Modell	LogLik (df)	AIC	BIC	LR-Test
1-PL	-4509.19 (452)	9140	9399	–
2-PL	-4458.83 (393)	9158	9666	$\Delta_{\chi^2}(59) = 100.7^{***}$

Anm.: $***p < .001$.

grundvariablen als etabliertes Verfahren identifizierbar ist (vgl. Abschnitt 4.4.2), wird es in die Berechnungen der vorliegenden Studie ebenfalls mit einbezogen. Tabelle 5.18 zeigt zunächst die bessere Passung eines eindimensionalen 2-PL-Modells mit den ausgewählten Hintergrundvariablen (vgl. Abschnitte 4.2.3 und 5.2.2). Bei größerer Modellsparsamkeit erklärt die Skalierung mit Hintergrundmodell die Daten im Vergleich besser. Die latenten Korrelationen zeigt Tabelle 5.19. Über eine theoretische, allgemeine, und nicht

Tabelle 5.18.: Relativer Modellvergleich zwischen eindimensionalem 2-PL-Modell mit und ohne Hintergrundvariablen der Modellaufgaben.

Modell	LogLik (df)	AIC	BIC	LR-Test
ohne Hintergrundmodell	-4458.83 (393)	9158	9666	–
mit Hintergrundmodell	-4394.16 (390)	9034	9556	$\Delta\chi^2(3) = 129.3^{***}$

Anm.: *** $p < .001$.

weiter differenzierte Modellnutzungsfähigkeit hinweg, haben alle Hintergrundvariablen einen Einfluss auf die Lösungswahrscheinlichkeit und können als bedeutsam beurteilt werden.

Tabelle 5.19.: Latente Korrelationen der Hintergrundvariablen für die eindimensionale Skalierung.

Variable	Koeffizient
kognitive Fähigkeiten	.32
Leseverständnis	.45
Fachwissen	.32

Anm.: Varianzaufklärung $R^2 = .44$

Multidimensionale Modelle

Tabelle 5.20 zeigt die Modellvergleiche unter verschiedenen, theoretisch plausiblen Strukturannahmen. Allen liegt die freie Schätzung der Steigungsparameter (2-PL) zugrunde. Im Überblick passen komplexere Modelle auf Kosten der Modellsparsamkeit signifikant besser auf die Daten als eine eindimensionale Lösung. Die Annahme einer Within-Item-Dimensionality führt in allen komplexen Strukturen zur besseren Passung. Im Zusammenspiel mit den theoretischen Überlegungen und den qualitativen Vorarbeiten, wird eine Einfachstruktur der Dimensionen darum verworfen.

Multidimensionale Modelle mit Hintergrundvariablen

Weil der Einfluss der Hintergrundvariablen deutlich ist, wurden die Strukturmodelle mit Within-Item-Dimensionality noch einmal unter Nutzung latenter Korrelationen spezifiziert und einem Vergleich unterzogen. Alle Modelle erklären die Daten bei geringerer Modellsparsamkeit besser als der 1-dimensionale Fall (Tabelle 5.21). Das 7-dimensionale Modell zeigt dabei auch den niedrigsten Wert im AIC. Dabei muss der Vergleich an

Tabelle 5.20.: Relative Modellvergleiche zwischen den verschiedenen Strukturmodellen ohne Hintergrundmodell.

Modell	LogLik (df)	AIC	BIC	LR-Test
eindimensional	-4458.83 (393)	9158	9666	–
3 Denkschritte (between)	-4457.24 (390)	9160	9682	$\Delta_{\chi^2}(3) = 3.2$
4 Modellkomponenten (between)	-4460.00 (387)	9172	9706	$\Delta_{\chi^2}(6) = -2.3$
7 Dimensionen (between)	-4398.85 (312)	9200	10052	$\Delta_{\chi^2}(81) = 119.9^{**}$
3 Denkschritte (within)	-4407.43 (333)	9175	9938	$\Delta_{\chi^2}(60) = 102.8^{***}$
4 Modellkomponenten (within)	-4383.41 (303)	9187	10077	$\Delta_{\chi^2}(90) = 150.8^{***}$
7 Dimensionen (within)	-4277.98 (183)	9216	10615	$\Delta_{\chi^2}(210) = 361.7^{***}$

Anm.: $^{**}p < .01$, $^{***}p < .001$.

Informationskriterien noch einmal gesondert vorgenommen werden (s. u.). Insbesondere werden folgend nur der korrigierte AIC (AICc) und der stichprobenadjustierte BIC (saBIC) als Vergleichsmaße herangezogen, weil sie für große Stichproben zu ihren nicht-korrigierten Analoga konvergieren (Henson et al., 2007) und bei kleineren Stichproben den Bias für das Anpassen der Daten an das Modell vermeiden (Burnham & Anderson, 2004). Weil das 7-dimensionale Modell beim AICc mit so deutlichen Straftermen für die

Tabelle 5.21.: Relative Modellvergleiche zwischen den verschiedenen Strukturmodellen mit Hintergrundmodell und Skalierung mit Within-Item-Dimensionality.

Modell	LogLik (df)	AIC	BIC	LR-Test
eindimensional	-4394.16 (390)	9034	9556	–
3 Denkschritte	-4341.17 (324)	9060	9862	$\Delta_{\chi^2}(66) = 105.9^{**}$
4 Modellkomponenten	-4304.89 (291)	9049	9990	$\Delta_{\chi^2}(99) = 183.4^{***}$
7 Dimensionen	-4100.31 (162)	8903	10391	$\Delta_{\chi^2}(228) = 587.7^{***}$

Anm.: $^{**}p < .01$, $^{***}p < .001$.

erhöhte Komplexität belegt wird, wird es für die folgenden Ausführungen, sowie die Diskussion nicht mehr berücksichtigt (vgl. Abschnitt 4.4.1). Das Verhältnis aus Parametern und Fallzahl ist so klein, dass es zugunsten eines konservativen Ausschlussverfahrens zurückgewiesen wird (Burnham & Anderson, 2004; Henson et al., 2007; Lorah & Womack, 2019). Das AICc steigt in dieser Studie mit erhöhter Modellkomplexität und liefert insofern ein kohärentes Bild (Tabelle 5.22). Um das 3- und 4-dimensionale Modell gegen Zufall abzusichern, wurden jeweils 10 Modelle spezifiziert, bei denen die Faktorstruktur gleich war, die Aufgaben diesen jedoch randomisiert zugewiesen wurden. Die erhaltenen Informationskriterien wurden nach Burnham und Anderson (2004) miteinander verglichen (Tabelle 5.23). Die berechneten Modellgewichte (vgl. Abschnitt 4.4.3) bevorzugen das 4-dimensionale Modell mit einer Wahrscheinlichkeit $w_i = .72$ deutlich und

das 3-dimensionale Modell verliert mit einer Wahrscheinlichkeit $w_i < .01$ gegenüber den Zufallsmodellen an Bedeutung. Weil unter Berücksichtigung der Modellvergleiche und

Tabelle 5.22.: Vergleich der Informationskriterien der verschiedenen Strukturmodellen mit Hintergrundmodell und unter Skalierung mit Within-Item-Dimensionality.

Modell	AIC	AICc	BIC	saBIC
eindimensional	9034.322	9112.739	9555.876	9164.495
3 Denkschritte	9060.330	9282.683	9861.743	9260.352
4 Modellkomponenten	9048.938	9390.359	9990.280	9283.884
7 Dimensionen	8902.626	10437.434	10390.963	9274.095

Anm.: Informationskriterien mit dem je niedrigsten Wert sind fett markiert.

der Informationskriterien das 4-dimensionale Modelle die beste Erklärung für die Daten darstellt, werden in Tabelle 5.24 die latenten Korrelationen der Modellnutzungsdimensionen mit den Variablen aus dem Hintergrundmodell berichtet. Auffällig sind die negativen Korrelationen in der Dimension *Operationen*. Tabelle 5.25 zeigt die EAP-Reliabilitäten der Dimensionen. Während die *kleinsten Sinneinheiten* und *Regeln* akzeptable Werte zeigen, müssen die Zahlen für *Relationen* als schwach reliabel und *Operationen* als inakzeptabel bewertet werden. Im Zusammenspiel mit den latenten Korrelationen muss die Deutung dieser Befunde besonders gründlich diskutiert werden. Die nicht weiter differenzierte Annahme der hierarchischen Schachtelung der Modellkomponenten im Sinne der theoretischen Struktur kann so nicht aufrechterhalten werden (vgl. Abschnitte 5.4 und 6.1.3).

5.3.4. Vergleich der Aufgabenschwierigkeiten

Die Aufgabenschwierigkeiten werden hier exemplarisch über das als am passendsten bewerteten 4-dimensionalen Modell inkl. Hintergrundvariablen über die Themenbezüge hinweg miteinander verglichen. In der Tendenz müssen, über alle Strukturannahmen hinweg, die Aufgaben zu den Strukturformeln als bedeutsam schwerer beurteilt werden. Dies wird auch über die multidimensionalen Schwierigkeiten (vgl. Abschnitt 4.4.2) aus der Schätzung ohne Hintergrundmodell gestützt. Zwischen den Aufgaben, die eine bestimmte Modellkomponente explizit fokussieren, sind keine bedeutsamen Schwierigkeitsdifferenzen festzustellen ($H(3) = 4.43, p > .05$). Anhang E zeigt diese nicht signifikanten Schwierigkeitsunterschiede grafisch, sowie diejenigen zwischen den drei Schritten des naturwissenschaftlichen Denkens ($H(2) = 1.36, p > .05$). Anhang B enthält die erhaltenen Aufgabenschwierigkeiten aller Aufgaben für die jeweiligen statistischen Modelle. Obwohl Varianzhomogenität nach Levene gegeben ist ($F(4, 55) = .599, p > .05$) und – sowohl über alle Themen hinweg als auch innerhalb – Normalverteilung nach Shapiro-Wilk angenommen werden kann ($W > .915, p > .05$), werden die die Schwierigkeiten konservativ mit dem nicht-parametrischen Kruskal-Wallis-Test vorgenommen. Der Grund dafür ist

Tabelle 5.23.: Vergleich des 3- und 4-dimensionalen Theoriemodells mit Zufallsmodellen anhand der Informationskriterien AICc und saBIC.

Modell	AICc	saBIC
3-Dim (Theorie)*	9277.358	9255.026
3-Dim (Zufall 1)	9297.920	9275.588
3-Dim (Zufall 2)	9276.500	9254.169
3-Dim (Zufall 3)	9298.544	9276.212
3-Dim (Zufall 4)	9278.863	9256.532
3-Dim (Zufall 5)	9270.355	9248.024
3-Dim (Zufall 6)	9282.351	9260.020
3-Dim (Zufall 7)	9267.990	9245.659
3-Dim (Zufall 8)	9288.027	9265.696
3-Dim (Zufall 9)	9264.383	9242.051
3-Dim (Zufall 10)	9280.649	9258.318
4-Dim (Theorie)**	9390.359	9283.884
4-Dim (Zufall 1)	9395.892	9289.417
4-Dim (Zufall 2)	9428.257	9321.782
4-Dim (Zufall 3)	9399.764	9293.290
4-Dim (Zufall 4)	9401.901	9295.426
4-Dim (Zufall 5)	9396.274	9289.799
4-Dim (Zufall 6)	9403.285	9296.811
4-Dim (Zufall 7)	9404.487	9298.013
4-Dim (Zufall 8)	9393.086	9286.611
4-Dim (Zufall 9)	9422.573	9316.098
4-Dim (Zufall 10)	9405.316	9298.841

Anm.: Die niedrigsten Werte sind fett gedruckt. *Wahrscheinlichkeit der besten Passung nach AICc und saBIC < 1%. **Wahrscheinlichkeit der besten Passung nach AICc und saBIC $\approx 72\%$.

Tabelle 5.24.: Latente Korrelationen der Modellnutzungsfähigkeit über die Faktoren der 4-dimensionalen Skalierung.

Dimension	LGVT	Fachwissen	KFT	R^2
kl. Sinneinheiten	.52	.40	.35	.52
Relationen	.22	.04	.09	.08
Operationen	-.28	-.10	-.09	.13
Regeln	.49	.31	.37	.48

Anm.: –

Tabelle 5.25.: EAP-Reliabilitäten der Faktoren aus der 4-dimensionalen Skalierung mit Hintergrundmodell.

Dimension	EAP-Reliabilität
kl. Sinneinheiten	.74
Relationen	.57
Operationen	.47
Regeln	.61

Anm.: –

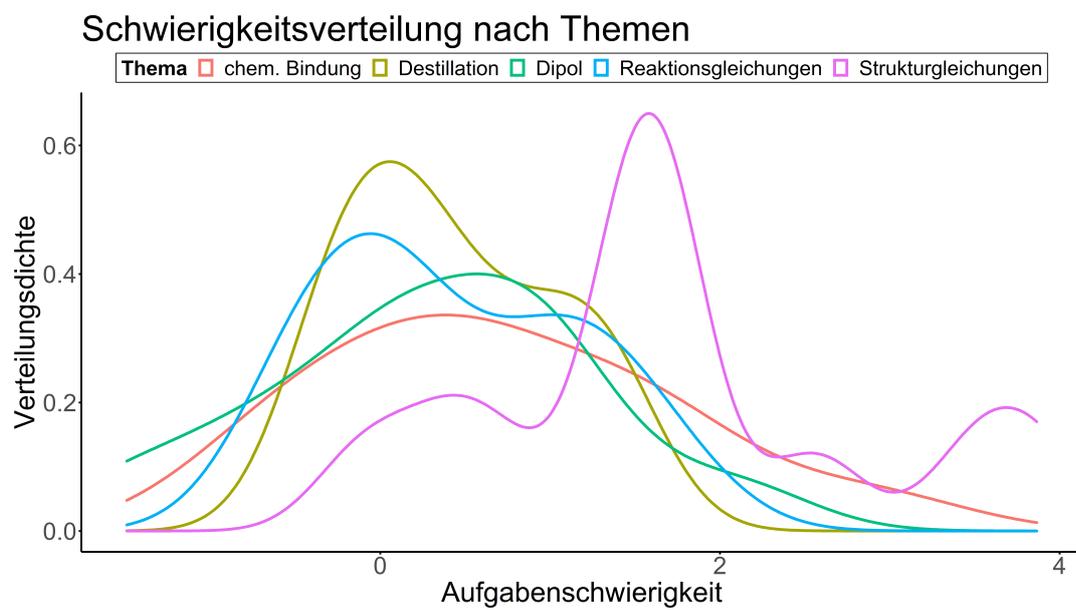


Abbildung 5.6.: Schwierigkeitsverteilung aus der 4-dimensionalen Aufgabenskalierung nach Themen.

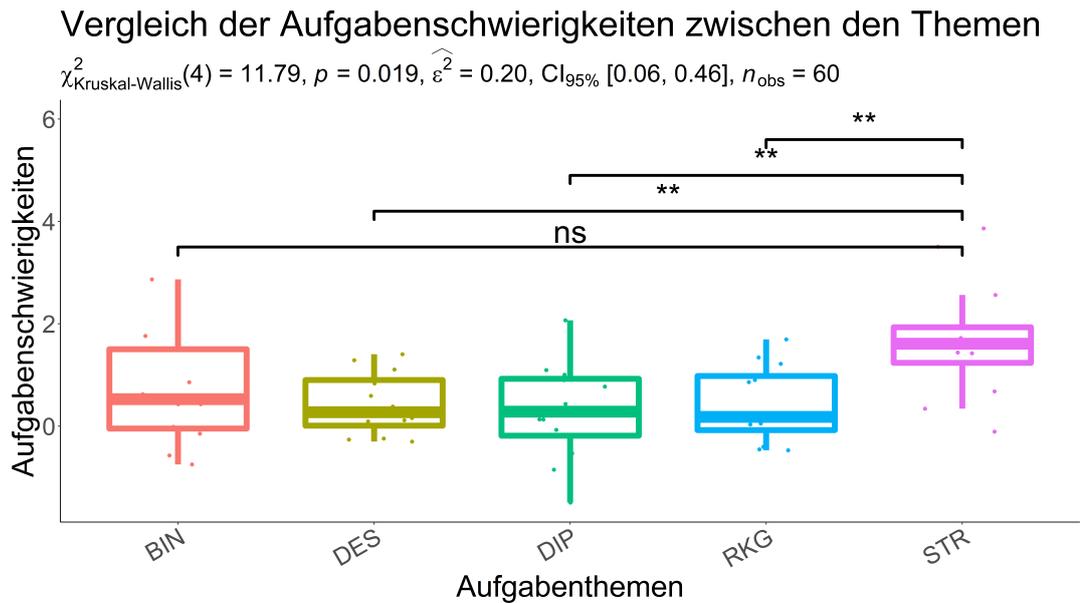


Abbildung 5.7.: Schwierigkeitsvergleiche aus der 4-dimensionalen Aufgabenskalierung nach Themen.

die verhältnismäßig geringe Anzahl Aufgaben pro Thema und die grafische Prüfung, die deutliche Unregelmäßigkeiten zeigt (Abb. 5.6). Der Test zeigt signifikante Unterschiede ($H(4) = 11.793, p < .05, \epsilon^2 = .2$) mit mittlerer Effektstärke. Die Post-hoc-Analyse nach Dunn zeigt nach Benjamini-Hochberg-Korrektur signifikante Paardifferenzen ($p < .05$) für drei von zehn Vergleichen. Davon sind immer Aufgaben mit Strukturformeln betroffen (Abb. 5.7). Weil für den Einfluss von Testhefteeffekten kontrolliert werden sollte, diese aber nicht Teil der Modellierung waren, wurde der Vergleich der Testheftschwierigkeiten über die Rohpunktwerte der Modellaufgaben vorgenommen. Dieser Vergleich (über alle Testhefttypen und innerhalb nicht normalverteilt nach Shapiro-Wilk: $W > .94, p < .01$, varianzhomogen nach Levene: $F(2, 510) = 2.47, p > .05$) zeigt keine signifikanten Unterschiede zwischen den Testhefttypen nach Kruskal-Wallis ($H(2) = 1.95, p > .05$). Tabelle 5.26 zeigt den deskriptiven Punktvergleich zwischen den Testhefttypen.

Tabelle 5.26.: Deskriptive Darstellung der erreichten Rohpunkte aus den Modellaufgaben nach Testhefttypen.

Testheft	$N_{\text{bearbeitet}}$	M (SD)	Median	SE
Random	185	5.6 (2.8)	5	.21
Denkschritt	111	5.9 (2.8)	6	.27
Modellkomponente	217	5.4 (2.4)	5	.16

Anm.: –

5.4. Zusammenfassung und Beantwortung der Forschungsfragen

Die Ergebnisse der empirischen Untersuchungen lassen sich überblicksweise wie folgt zusammenfassen:

- Per Auswertung des Expertenratings und der offenen Aufgaben bei SuS, war eine klare Distinktion der verschiedenen Modellkomponenten als einzeln messbare Dimensionen der Modellnutzung empirisch nicht möglich und im Sinne des nomologischen Netzes der Arbeit auch theoretisch nicht gerechtfertigt. Die Experten äußerten explizit theoriebezogene Zweifel an solch einer Trennbarkeit. Die SuS gaben bei Fragen nach Relationen, Operationen und Regeln grundsätzlich immer mindestens auch Modellbestandteile an, die als kleinste Sinneinheiten interpretiert werden konnten. Diese Befunde stützten die Annahme einer hierarchischen Struktur bei der Modellnutzung, bei der jene kleinsten Sinneinheiten als basale Grundbestandteile immer enthalten sind.
- Die quantitativen Vorarbeiten mit Hilfe von IRT-Modellen, implizierten eine grundsätzliche Messbarmachung für die Fähigkeit die konstruktbezogenen Aufgaben lösen zu können. Die Modellannahmen eines eindimensionalen Konstrukts konnten aber aufgrund der ICCs und der DETECT-Statistik nicht uneingeschränkt beibehalten werden, auch wenn die Fit-Werte aus der 2-PL-Skalierung eine gute Passung implizierten.
- Die Korrelationen der Modellnutzungsfähigkeit mit den Kovariablen KFT, Fachwissen und LGVT waren in der Vorstudie statistisch bedeutsam und beeinflussten die Lösungswahrscheinlichkeit für die Modellnutzungsaufgaben auch über die Gymnasien unterschiedlich stark. Aus diesem Grund wurden sie als Variablen per Hintergrundmodell direkt in die Schätzungen der Hauptstudie integriert. Auf diese Weise sollte für ihren Einfluss kontrolliert werden.
- In der Hauptstudie wurden mehrdimensionale Modelle sowohl mit Between-Item-Dimensionality, als auch mit Within-Item-Dimensionality inkl. Hintergrundmodell spezifiziert und miteinander verglichen. Unter Berücksichtigung von Modellsparamkeitskriterien auch gegenüber Zufallsmodellen und den theoretischen Strukturannahmen, wird das 4-dimensionale Modell mit hierarchisch geschachtelten Modellkomponenten als am besten passend bewertet. Anhang B dokumentiert die Steigungsparameter der jeweiligen Skalierungen, über die die Dimensionalitätszuordnungen quantifiziert wurden.

5.4.1. Beantwortung der Forschungsfrage I

Die Frage *Entspricht die empirische Struktur der Testergebnisse der theoretisch angenommenen Struktur der hypothetisch-deduktiven Modellnutzung im Fach Chemie?* kann

5. Ergebnisse

unter Rückgriff auf die korrespondierenden Forschungshypothesen eingeschränkt positiv beantwortet werden.

Zur Hypothese Ia

Die Forschungshypothese H_1 lautete: *Ein 7-dimensionales Modell passt signifikant besser auf die Testdaten, als ein 4-, 3-, oder 1-dimensionales Modell.*

Nach den statistischen Tests wird die Nullhypothese H_0 (*Höherdimensionale Modelle lassen sich statistisch nicht von Modellen mit geringerer Dimensionalität unterscheiden.*) zurückgewiesen.

Obwohl die Forschungshypothese statistisch beibehalten werden kann, muss sie unter den gegebenen Befunden weiter eingeschränkt werden. Folgende Aussagen treffen nur auf die Skalierungen mit Hintergrundmodell und Within-Item-Dimensionality zu (s. u. zur Hypothese 1b). Das Kriterium der Modellsparsamkeit zwingt zur Zurückweisung des 7-dimensionalen Modells, weil der Kennwert des AICc um mehr als eine Größenordnung höher liegt, als für die Vergleichsmodelle. Bei der weiteren Abwägung zwischen dem verbliebenen 3- und 4-dimensionalen Konstrukt per AICc, wurde letzterem der Vorzug gegeben, ohne ersteres vollständig zurückweisen zu können. Der Messung liegt die Annahme zugrunde, dass die hierarchisch jeweils niedrigste Einheit (*kleinste Sinneinheit* bei vier Dimensionen, bzw. *Fragestellung/Hypothese* bei drei Dimensionen) in jeder Aufgabe vorhanden ist. Bei der zufälligen Zuordnung von Aufgaben in die jeweiligen Kategorien für die Zufallsabsicherung, kommt es also zwangsläufig zu Unschärfen, weil das confirmatorische Modell gezwungen wird, diese Aufgaben in jedem Fall als korrekt zugeordnet anzunehmen (vgl. Abschnitt 6.1.1).

Zur Hypothese Ib

Die Forschungshypothese H_1 lautete: *Ein Modell mit einer Within-Item-Dimensionality passt signifikant besser auf die Testdaten, als ein Modell mit einer Einfachstruktur.*

Nach den statistischen Tests wird die Nullhypothese H_0 (*Ein Modell mit einer Einfachstruktur und ein Modell mit Within-Item-Dimensionality lassen sich statistisch nicht voneinander unterscheiden.*) zurückgewiesen.

Während die mehrdimensionalen Skalierungen mit Einfachstruktur die Daten nicht besser erklären als der eindimensionale Fall, trifft dies auf die verschachtelten Strukturannahmen nicht zu. Sie passen besser, können theoretisch gerechtfertigt werden und erlauben eine differenzierte Fähigkeitsmessung. Darüber hinaus muss berücksichtigt werden, dass das eindimensionale 2-PL-Modell zwar als restriktives Referenzmodell für die Vergleiche benutzt wurde und die größte Modellsparsamkeit aufweist, unter Berücksichtigung der Steigungen in den ICCs allerdings gar nicht vollumfänglich als gültig beurteilt werden kann.

5.4.2. Beantwortung der Forschungsfrage II

Die Frage *Ist die Modellnutzung in der dargestellten Form unabhängig von den jeweiligen Themenbezügen?* muss unter Rückgriff auf die korrespondierende Forschungshypothese verneint werden.

Zur Hypothese IIa

Die Forschungshypothese H_1 lautete: *Die mittleren Aufgabenschwierigkeiten aus den IRT-Skalierungen weichen zwischen den Aufgabenthemen signifikant voneinander ab.*

Nach den statistischen Tests wird die Nullhypothese H_0 (*Zwischen den Aufgabenschwierigkeiten der jeweiligen Themen gibt es keine signifikante Abweichung.*) zurückgewiesen.

Zur Konstruktion des Tests gehörte die Annahme, dass die Aufgabenschwierigkeiten sich über alle Themenbereiche hinweg ungefähr gleich verteilen. Hätte die Nullhypothese nicht zurückgewiesen werden können, so hätte diese Annahme bestehen bleiben können. Da dies nicht der Fall ist, muss der statistisch signifikante Schwierigkeitsunterschied diskutiert und sein Einfluss, vor allem auf die curriculare und interne Validität, ins Verhältnis zu den anderen Befunden gesetzt werden.

6. Diskussion

Das folgende Kapitel greift die empirischen Befunde der vorliegenden Arbeit auf, bezieht sie aufeinander und führt sie mit dem Theorieteil zusammen. Dabei werden die Ergebnisse auch im Bezug auf die aktuelle Forschungslage eingeordnet. Der Interpretationsraum für die Befunde ist hinreichend groß, sodass weiterführende Literatur aus anderen Forschungsbereichen bei der Bewertung eingesetzt werden. Auf diese Weise soll der Horizont der Arbeit, über die Schwerpunkttheorien hinaus, erweitert werden. Weil der Modellbegriff nicht nur in den erkenntnistheoretischen Grundannahmen, sondern auch als statistischer Grundbegriff eine wichtige Rolle in der Arbeit spielt, wird eine reflexive Wendung der Modelltheorie auf die Methoden vorgenommen. Abschließend werden die diskutierten Punkte noch einmal überblicksweise dargestellt und ein Ausblick auf mögliche Anschlussforschung gegeben.

6.1. Interpretationen der Befunde

Das primäre Anliegen der Arbeit war die hypothesenbasierte Strukturaufklärung von kompetenzorientiert konstruierten Leistungsaufgaben (Kapitel 2). Die Antwortmuster wurden statistischen Tests unterzogen (Kapitel 4), die zu numerischen Ergebnissen führten (Kapitel 5). Auf dieser Basis konnten bestimmte Annahmen aufrechterhalten oder zurückgewiesen werden (Kapitel 3). Dies führt im Folgenden zu den Deutungen über die strukturelle Beschaffenheit der Modellnutzung.

6.1.1. Dimensionalität des Konstrukts

Für die kritische Diskussion der Dimensionalitätsuntersuchung ist es geboten, alle Skalierungen der Testaufgaben noch einmal zu revidieren und die Aspekte ihrer Gültigkeit einander gegenüberzustellen. Grundsätzlich gilt bei diesen Vergleichen das Sparsamkeitsprinzip als Leitmotiv. So kommt der 1-dimensionalen Testskalierung ohne Hintergrundmodell eine besondere Bedeutung zu. Es handelt sich dabei um die einfachste Strukturannahme für die Beschreibung der Modellnutzungsfähigkeit (Abschnitt 2.4.4), die von einer übergeordneten und nicht weiter differenzierbaren, aber durchaus messbaren Fähigkeit ausgeht. Dabei wäre es nicht überraschend gewesen, wenn das entsprechende Messmodell die Daten gut passend erklärt hätte, obwohl die Items sensitiv für Dimensionsunterschiede sind (vgl. den 1-dimensionalen Strukturvorschlag in Abschnitt 2.4.4 sowie Reckase, 2009, S. 194 ff.). Neben der nachweisbaren Abgrenzung von kognitiven Fähigkeiten, Leseverständnis und Fachwissen, sowie der messproduktiven Fit-Kennwerte

(Abschnitt 5.2.2), spielten die ICCs eine wichtige Rolle. Die Interpretation dieser Modellkurven lautete für einige Aufgaben: Die Lösungswahrscheinlichkeit sinkt mit steigender Personenfähigkeit. Im Sinne der Validitätsüberlegungen (vgl. Abschnitt 4.2.1) konnte eine Eindimensionalität daher nicht als zufriedenstellende Strukturannahme aufrechterhalten werden. Nichtsdestotrotz ist es möglich, das eindimensionale Modell als Schätzung für Fähigkeitsunterschiede, bspw. zwischen Schulen zu verwenden (Camilli & Fox, 2015; Crişan et al., 2017).

Wie in den Abschnitten 4.4.1 und 5.3.3 beschrieben, können die Steigungsparameter von IRT-skalierten Aufgaben als Analoga von Faktorladungen aus konfirmatorischen Faktoranalysen interpretiert werden (Bühner, 2011; Chalmers, 2012). Auf diese Weise ist eine entsprechende Dimensionalitätszuordnung quantifizierbar. Das wurde in der vorliegenden Studie angewendet. Dabei wurde – unter Berücksichtigung der Modellpassung – die Skalierung von vier Within-Item-Dimensionen mit Hintergrundmodell als plausibelste Erklärung für die gefundenen Antwortmuster identifiziert (Abschnitt 5.3.3). Ausschlaggebend war dabei der Vergleich mit Zufallsmodellen, in denen sich der parallele 3-dimensionale Fall nicht gegen Zufallsmodelle durchsetzen konnte. Dieser Befund muss allerdings eingeschränkt werden, weil der Vergleich über Modellgewichte im Fall der ineinander geschachtelten Dimensionen mit Unschärfen behaftet ist. Da die Zuordnung zu *Frage/Hypothese* (3-dimensionaler Fall) hier in jeder Aufgabe gegeben ist, muss angenommen werden, dass die berechnete, bessere Passung durch ein Zufallsmodell gar nicht vollständig zufällig ist, sondern durchaus die theoriekonformen Zuordnungen widerspiegelt. Insofern kann auch die Modellierung einer Fähigkeit zum Durchlaufen des hypothetisch-deduktiven Erkenntniswegs *an Modellen* als valide Alternative für die Erklärung der Leistungsunterschiede benannt werden (vgl. Abschnitt 2.4.4). Dies steht im Einklang mit Ergebnissen aus quantitativen (Grünkorn, 2014; Nehring et al., 2015) und qualitativen (Moreira et al., 2019; Oliva et al., 2015) Studien zur Untersuchung naturwissenschaftlicher Denkprozesse. Das kann auch aus konzeptuellen (Gilbert & Justi, 2016) und kompetenzorientierten (Upmeyer zu Belzen & Krüger, 2010) Überblicksarbeiten zum prozessorientierten Charakter der Modellnutzung abgeleitet werden.

6.1.2. Aufgabenschwierigkeiten

Während die Annahme distinkter Fähigkeitsdimensionen, d. h. eine Einfachstruktur, aufgrund der Modellvergleiche für eine Strukturaufklärung zurückgewiesen werden musste, konnten die entsprechenden Skalierungen über die multidimensionalen Schwierigkeitsschätzungen Hinweise auf konvergente Validität liefern: Obwohl solch eine Strukturannahme theoretisch und empirisch nicht vollständig haltbar ist, zeigen die Schwierigkeitsparameter der Aufgaben für alle berechneten Modelle bei gleicher Größenordnung in die gleiche Richtung. Signifikante Unterschiede können nur zwischen den einzelnen Themenbereichen festgestellt werden (Abschnitt 5.3.4), weswegen von einem themenspezifischen, d. h. curricularen Einfluss ausgegangen werden kann. Ein möglicher Unterschied in der Aufgabenkomplexität, z. B. zwischen einer Aufgabe zu den *kleinsten Sinneinheiten* versus einer Aufgabe zu den sich darauf beziehenden *Relationen* (vgl. Woitkowski & Riese, 2017) scheint insofern ebenfalls keine bedeutsame Rolle zu spielen. Dies wird auf

die starke Standardisierung der Aufgabenstämme und Auswahlmöglichkeiten zurückgeführt. Reihenfolge- bzw. Ermüdungseffekte zu modellieren war nicht Ziel der Studie und der nicht signifikante Unterschied zwischen den erworbenen Punkten über die Testhefttypen lässt den Schluss zu, dass diese Effekte entweder keinen Einfluss hatten, oder ihre Kontrolle mithilfe des Testheftdesigns gelungen ist.

6.1.3. Einfluss der Kovariablen

Die Kovariablen wurden über ein Hintergrundmodell in die Fähigkeitsschätzungen eingeführt. Für die Dimensionen der *kleinsten Sinneinheiten* und der *Regeln* sind über die latenten Korrelationen deutliche Einflüsse der Kovariablen messbar, die in beiden Fällen ca. 50% der Varianz aufklären. Die negativen Koeffizienten der Hintergrundmodellvariablen mit der Dimension der *Operationen* kann aus verschiedenen Perspektiven diskutiert werden. Im Extremfall können derartige Effekte durch falsche Modellspezifikationen und große Messfehler zustande kommen (Bühner, 2011; Cole & Preacher, 2014). Für diese strenge Auslegung sprechen zunächst die vergleichsweise niedrigen EAP-Reliabilitäten der *Operationen* sowie der *Relationen* und deren niedrige Varianzaufklärung von 13, bzw. 8% durch die Kovariablen. Es ist nicht unmittelbar einsichtig, weswegen die Diskrepanz der Einflüsse derartig hoch sein sollte, wenn gleichzeitig die anderen empirischen Befunde für eine hohe Standardisierung bei der Aufgabenkonstruktion sprechen (vgl. Abschnitte 4.1, 4.2.2 sowie 5.1). Andererseits wird die theoretisch angenommene Struktur durch den Vergleich mit Zufallsmodellen explizit gestützt. Dieses gemischte Bild bedarf der Erweiterung der hypothetischen Struktur, weil die Befunde an dieser Stelle heterogen sind. Aus den Ergebnissen der Arbeit werden zwei mögliche Erklärungsmuster abgeleitet. Erstens konnten in der Vorstudie auch zwischen den Gymnasien signifikante Unterschiede in den Kovariablen nachgewiesen werden. Solche Schuleffekte sind für Tests mit Within-Item-Dimensionality innerhalb von Mehrebenenanalysen belegt (Zhang et al., 2019), konnten aber in der vorliegenden Arbeit aufgrund der begrenzten TeilnehmerInnenzahl nicht modelliert werden. Darüber hinaus kann, zweitens, eine inhaltliche Überlegung zur Struktur der Modellkomponenten angestellt werden, die die hohen Zusammenhänge mit zwei der Komponenten einerseits, und den niedrigen Zusammenhängen mit den beiden anderen Komponenten andererseits, erklärt: Sie fallen in eine andere Struktur der Modellnutzung. Die *Sinneinheiten* und die *Regeln* können, nach der Revision der erhaltenen Ergebnisse, in ihrer Beschaffenheit als hypothetische Zustandsbeschreibungen charakterisiert werden. In diesem Fall würden sie genutzt, um zu beschreiben was für ein Sachverhalt vorliegt und nicht, wie er sich verändert. Die *Relationen* und im besonderen Maße die *Operationen*, die sich aus der Theorie heraus auf die Veränderungen zwischen den Relationen beziehen, können als stärker dynamisch und mit größerem prädiktivem Charakter beschrieben werden (für eine kognitionspsychologische Perspektive vgl. Simms & Richland, 2019). Gerade der Fokus auf Veränderungen impliziert Bewegungen in den Komponenten und bedarf möglicherweise einer anderen kognitiven Subdimension bei der Modellnutzung. Dies kann auch an den theoretischen Vorarbeiten der Studie kontrastiert werden. In den fünf Charakteristika für Modellnutzung (vgl. Abschnitt 2.2.1) nach Knuuttila (2011) hat die *konkrete Manipulierbarkeit* in empirischer Hinsicht großes Re-

flexionspotential, weil sie sich auf eine konkrete, empirische Handlung im Rahmen eines Erkenntnisprozesses bezieht. Wenn das zutreffen würde, dann wäre die Spezifikation mit einer 2-dimensionalen Einfachstruktur passender. Die empirisch gefundenen Kovarianzen, die aufgrund der angenommenen Within-Item-Dimensionality zustande kommen, würde in diesem Fall von einer sparsameren Strukturannahme dominiert. Erste Hinweise darauf liefern Rost und Tiemann (2020) unter Spezifikation eines G-Faktor-Modells. Das wird als ein Desiderat dieser Studie konstatiert.

6.1.4. Rückbindung an die epistemologischen Vorüberlegungen

In der vorliegenden Arbeit wird Erkenntnisgewinnung als hypothetisch-deduktives Verfahren definiert. Dies stellt aber sowohl erkenntnistheoretisch als auch fachdidaktisch nur eine Definitionsmöglichkeit dar. Heering und Kremer (2018) zeigen deutlich, dass Erkenntnisgewinnung nicht nur anders definiert werden *kann*, sondern vor dem breiten Horizont der *Natur der Naturwissenschaften* deutlich mehr sein *muss*, als ein mechanisch abzuarbeitendes Verfahren. Dabei ist es auch nicht relevant, ob deduktive oder induktive Verfahren zum Einsatz kommen. In der Konsequenz ist es nötig, dass an dieser Stelle eine deutliche Einschränkung der Verallgemeinerbarkeit der Studienergebnisse betont wird: Modellnutzung – im dargestellten und untersuchten Sinne – ist ein wichtiger Teil der Naturwissenschaften und des naturwissenschaftlichen Unterrichts. Ist die Modellnutzung als streng hypothetisch-deduktiv definierter Prozess definiert, kann außerdem eine Messbarkeit gewährleistet werden. Aspekte wie Kreativität, Auseinandersetzung mit dem Verhältnis von Selbst und Gegenstand, oder gar eine Einbindung in so etwas wie eine naturwissenschaftliche Gemeinschaft, sowie das damit verbundene kollaborative Nachdenken und Austausch, sind aber vollständig ausgeblendet. Verbindungen zu diesen Bestandteilen können folglich nur mittelbar und post-hoc skizziert werden. Solche Überlegungen sind zwar in den erkenntnistheoretischen Grundlagen der Arbeit angelegt, meist aber auch dort nicht unmittelbar zu finden. Der *Modellcargo* von Mahr (2008, 2011, 2015) kann als Kommunikationsmittel mit anderen Akteuren interpretiert werden. Expliziert wird dies in den entsprechenden Texten aber nicht. Van Fraassen (1980) verweist zwar immer wieder auf das Akzeptanzkriterium für naturwissenschaftliche Theorien, setzt diesem aber keinen expliziten, funktionalen Mechanismus voraus. Während Devitt (2006) und Knuuttila (2011) zwar eine scharfe Analyse der Widersprüche in häufig genutzten, modellbasierten Erkenntnismethoden vornehmen, setzen sie implizit stets eine Intersubjektivität voraus, die nicht weiter begründet wird und die sich auch bei Popper (1935) wiederfindet. Er behauptet Objektivität über Satzketten (aus denen Theorien bestehen) (ebd., S. 60 ff.), deren Zustimmungskriterien durch Gemeinschaften aus dem Text heraus gar nicht bestimmbar sind. Möglicherweise sollten soziologische Arbeiten in dieser Hinsicht als zusätzliche Untersuchungsfacetten für fachdidaktische Forschung untersucht werden. Dies könnte solipsistische Erklärungsmuster aus der Epistemologie deutlich erweitern. Aus einer ähnlichen Perspektive schlagen bspw. Oliva et al. (2015) einen sozio-konstruktivistischen Ansatz vor. Um aber stärker anschlussfähig an bisherige Überlegungen zu bleiben, wird die Aufmerksamkeit hier eher zu sprachphilosophischen und argumentationstheoretischen Theorien gelenkt. Die entsprechende Untersuchung der an-

gesprochenen Satzketten scheint aufgrund der binären Wahr-Falsch-Struktur erfolgsversprechend zu sein. Entsprechende Forschungslücken wurden bereits angedeutet (vgl. Abschnitt 2.1.1), sind aber noch nicht theoretisch und empirisch konkretisiert. Die zusammenfassende These lautet, dass Sprache als Verbindung, sowohl zwischen den Gedanken eines Subjekts, als auch zwischen den Gedanken mehrerer Akteure, im Kontext naturwissenschaftlicher Erkenntnisgewinnung noch zu wenig systematische Aufmerksamkeit erhält. Obwohl Sprache als konstituierendes Merkmal des Lernens einen selbstverständlichen Platz in der Forschungslandschaft besitzt (Aduriz-Bravo & Revel Chion, 2017; Carlsen, 2007; Espinet et al., 2012; Osborne, 2012), sind aktuelle Forschungsarbeiten dazu oft auf die Sichtebeine fokussiert. Behling et al. (2019) untersuchen bspw. den Umgang von Lehrkräften mit bildungssprachlichen Merkmalen im Biologieunterricht und Härtig et al. (2019) quantifizieren den Einfluss von Verständlichkeitsmerkmalen auf Fachwissenserwerb in Chemie und Physik. Insbesondere die Nutzung von Modellen wird allerdings nicht prominent mit Sprache in naturwissenschaftlichen Lehr-/Lernumgebungen zusammengeführt (vgl. die Übersichten in Gilbert & Justi, 2016; Gilbert & Treagust, 2009; Upmeyer zu Belzen et al., 2019). Für chemiespezifische Untersuchungen stehen in der Literatur außerdem kaum quantitative Untersuchungen zur Verfügung. Meist werden entsprechende Überlegungen qualitativ und wenig empirisch vorgenommen (Aduriz-Bravo, 2013; Laszlo, 2013).

Mit diesem Überblick und dem nicht gelösten Problem der verschiedenen Begriffsverwendungen (vgl. Kapitel 1, v. a. Nicolaou & Constantinou, 2014) kann pragmatisch umgegangen werden. Es besteht nicht notwendigerweise eine Generalisierbarkeitspflicht für fachdidaktische Forschungsprojekte (vgl. Schecker et al., 2014, S. 5) oder auch für empirische Sozialforschung im Allgemeinen (Amrhein et al., 2019). Das ist aber nicht nur aus der individuellen Sicht eines Forschers unbefriedigend, sondern kann ein Hinweis auf mangelnde Theoriegebäude sein. Beispielsweise kann die epistemologische Haltung – d. h. auch die Explikation derselben – von Lernenden in empirischer Hinsicht sowohl als einflussreich (bspw. auf das konzeptuelle Verständnis von Sachverhalten), als auch, speziell für das Fach Chemie, als zu wenig longitudinal und experimentell beforscht festgestellt werden (Yang & Tsai, 2012). In der Konklusion scheint es also durchaus angemessen, weiterhin nach verallgemeinerbaren Theoriestützcungen zu suchen – auch außerhalb etablierter fachdidaktischer Rahmungen. Es wird der Vorschlag gemacht, sich diesem Desiderat mit Überlegungen aus der analytischen Argumentationstheorie zu nähern (vgl. Abschnitt 6.2).

Reflexive Wendung der Modelltheorie auf die Methode

Weil das Vorgehen bei der Strukturprüfung der Modellnutzung hypothetisch-deduktiver Natur ist, können die Annahmen durch die Ergebnisse nur falsifiziert, aber niemals verifiziert werden. Die durchgeführten Hypothesentests sind streng auf der Grundlage der Zurückweisung der Nullhypothesen, bzw. der darunterliegenden Annahmen durchgeführt worden und machen keine Aussage darüber, ob die entgegengesetzten Strukturvorschläge korrekt sind. Auf diese Weise treffen sich die Einschränkungen solcher Tests und der entsprechenden Kennwerte (Head et al., 2015) mit rein qualitativ-wissenschaftstheoretischen

Beschreibungen naturwissenschaftlicher Erkenntnisgewinnung:

„Wir behaupten, daß die Fragen, die sich an den Begriff der Hypothesenwahrscheinlichkeit knüpfen, durch wahrscheinlichkeitslogische Überlegungen überhaupt nicht berührt werden: sagt man von einer Hypothese, sie sei nicht wahr, aber 'wahrscheinlich', so kann diese Aussage unter keinen Umständen in eine Aussage über eine Ereigniswahrscheinlichkeit umgeformt werden.“
(Popper, 1935)

Oder, nach Wasserstein und Lazar (2016): „No single index should substitute for scientific reasoning.“ Die vorliegende Arbeit kann den Messergebnissen in der Folge nur plausible Deutungsmöglichkeiten zuweisen (s. o.), nicht aber von statistischer Signifikanz auf eine Verifizierung der theoretischen Struktur schließen. Saint-Mont (2011) begründet diese Warnung in einer historischen Zusammenfassung und nennt die Fixierung auf Signifikanzniveaus und Hypothesentests mit mechanischer Suche nach p-Werten eine „kochbuchartige Schulstatistik“ (ebd. S. 93). In einer Analyse des LR-Tests (vgl. Abschnitt 4.4.3) zum Modellvergleich stellt er heraus, dass Forscher vor allem am *worst case* interessiert seien, das Unterschiedsmaß (d. h. der p-Wert) kein quantifizierbares Wahrscheinlichkeitsmaß darstelle und die Menge an Alternativhypothesen überhaupt nicht trivial zu bemessen sei (ebd., S. 95 ff.). Aus diesem Grund werden im Abschnitt 6.1.5 Folgeuntersuchungen vorgeschlagen, um Möglichkeiten zur Horizonterweiterung zu schaffen, „Strohänner“ als Nullhypothesen zu vermeiden und Alternativerklärungen anzubieten.

6.1.5. Vorschläge für unmittelbar mögliche Folgeuntersuchungen

Für einen schlagwortartigen Überblick der folgenden Vorschläge kann Abbildung 6.1 herangezogen werden. Der bereits genannte Zusammenhang mit empirisch geprüften Konstrukten wie dem Kompetenzmodell der Modellkompetenz (Upmeier zu Belzen & Krüger, 2010) ist durch weitere Anbindungen an fachdidaktisch-pädagogische Untersuchungen zu ergänzen. Die Ergebnisse der vorliegenden Arbeit zeigen einen bedeutsamen Einfluss von Fachwissen und einen curricularen, bzw. themenspezifischen Effekt. Fulmer et al. (2015) weisen darauf hin, dass die Aufgabenschwierigkeiten nicht vom Kontext einer jeweiligen Aufgabe trennbar sind und dass gleichzeitig Fachwissen und naturwissenschaftliche Denkprozesse analytisch voneinander unterscheidbar sind.

Eine andere Perspektive auf die Befunde kann vom Standpunkt der domänenspezifischen Metakognition (Andersen & Garcia-Mila, 2017; Cooper & Sandi-Urena, 2009; Cooper et al., 2008) aus eingenommen werden. Unter der Annahme, dass die beschriebenen Modellkomponenten keine intrinsischen, kognitiven Strukturen, sondern instruktive Scaffolds sind (vgl. Caspari, Kranz et al., 2018; Caspari, Weinrich et al., 2018), wären die Formulierungen in den konstruierten Aufgaben weniger als Rekonstruktion von gedanklichen Prozessen zu interpretieren. Sie könnten vielmehr als metakognitive Prompts aufgefasst werden, die beim situativen Nachdenken und Lösen der Aufgaben als Hilfsmittel dienen.

Bezogen auf das als hypothetisch-deduktiv modellierte, naturwissenschaftliche Denken in drei messbaren Schritten, können Überlegungen zur Deutung dieser Denkweise als

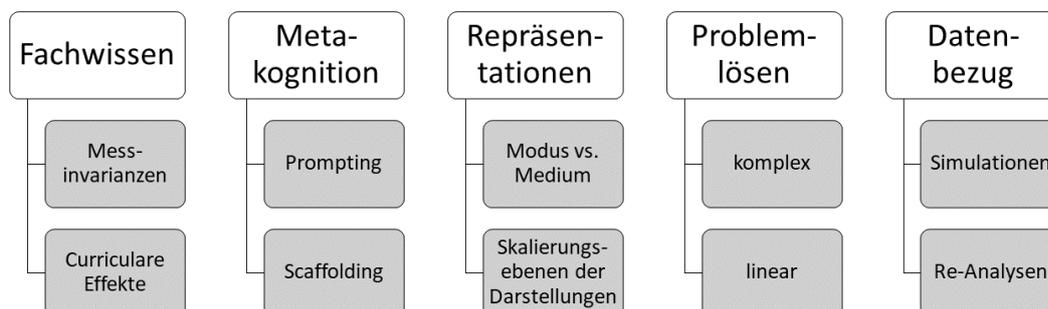


Abbildung 6.1.: Überblick der Facetten von möglichen Anschlussuntersuchungen.

lineares oder komplexes Problemlösen angestellt werden. Für chemiespezifisches, komplexes Problemlösen und dessen Zusammenhänge, bspw. mit Fachwissen, liegen bereits empirische Studien vor (Scherer & Beckmann, 2014). Darum wäre die Integration von modellbezogenen, problemhaltigen Umgebungen eine Möglichkeit, um die Fachspezifität noch um eine naturwissenschaftsübergreifende Facette zu erweitern.

Desweiteren wird hier auf die Anschlussfähigkeit zur Unterscheidung von Repräsentationsmodus und -medium hingewiesen (vgl. Knuuttila, 2011, in Abschnitt 2.2.1), die in Ergänzung zu Untersuchungen von Repräsentationen im Chemieunterricht (Dori & Karberman, 2012; Luxford & Bretz, 2013; Sumfleth & Nakoinz, 2019) herangezogen werden kann. Nur weil es in erkenntnistheoretischer Hinsicht nicht trivial ist, das Verhältnis von Modellen und Repräsentationen klar zu beschreiben, können Untersuchungsdesigns, die den Aspekt der Repräsentation in den Vordergrund stellen, trotzdem zielführend und erkenntniserweiternd sein.

Für alle vier exemplarisch genannten Ansätze können experimentelle und quasi experimentelle Umgebungen aufgesetzt werden, die entsprechende Zusammenhangshypothesen durch konvergente und/oder diskriminante Untersuchungsstrategien empirisch zugänglich machen. Cheng und Lin (2015) zeigen dies bspw. an den Einstellungen zu, bzw. dem Wissen über Modelle und weisen einen positiven Zusammenhang mit Lernerfolg nach. Ergänzend plädieren Erduran und Kaya (2019) für die Implementation von erkenntnistheoretischen Kernkonzepten in die (Chemie-)lehrerbildung und Danusso et al. (2010) demonstrieren die Wirksamkeit bei der konsequenten Einbettung von Wissen *über* Modelle in der Lehrerbildung.

Stärker datengetriebene Ansätze für Folgestudien wären Simulationsstudien. In diesen können Schul- oder Kontexteffekte berücksichtigt werden (Rose et al., 2019; Weirich, 2015) und/oder die große Missingquote in den Antworten unter Berücksichtigung

der Kovariablen imputiert werden (Glas et al., 2015). Konstruiert man dann mit entsprechend simulierten Datensätzen Bedingungen wie sie in den tatsächlichen Datenerhebungen stattgefunden haben, kann man die erhaltenen Schätzungen miteinander abgleichen. Auf diese Weise können dann zielführende Erhebungsstrategien für zukünftige Feldstudien abgeleitet werden. Diverse Autoren für Erweiterungspakete der Statistiksoftware *R* stellen dafür eine Reihe von Zugängen zur Verfügung, die sich explizit auf IRT-Modelle (Fox, 2010), oder auch auf die klassische Testtheorie (McElreath, 2020) beziehen, moderne Implementationen des Bayesianischen Ansatzes für die Parameterabschätzung einsetzen und gleichzeitig verhältnismäßig übersichtlich zu bedienen sind (Lee et al., 2020).

6.2. Zusammenfassung und Ausblick

Die hier erarbeiteten Ergebnisse zeigen den Forschungsbedarf am individuell-situativen Moment bei der Nutzung von Modellen (vgl. auch Mahr, 2015; Stachowiak, 1973; van Fraassen, 1980, aus Abschnitt 2.2). Dies wird aus der starken Abhängigkeit von Kovariablen, sowie der nicht hinreichenden Eindeutigkeit der Ergebnisse geschlossen. Zur Anschlussfähigkeit der Resultate an weiterführende Untersuchungen wird an dieser Stelle auf die Notwendigkeit eines triangulierenden Verfahrens hingewiesen, das weniger an der Strukturaufklärung einzelner Aufgaben oder Testinstrumente interessiert ist, sondern viel mehr die gedanklichen Prozesse der jeweiligen TeilnehmerInnen in den Fokus stellt. Für einen Überblick mit vielen Berührungspunkten sei auf Nitz und Fechner (2018) verwiesen, die diesen Bereich mit dem Begriff der *mentalen Modelle* abstecken. Johnson-Laird (2010) ergänzt diese mögliche Ausrichtung aus einer kognitionspsychologischen Perspektive. Mit den strukturellen Vorschlägen der Modellnutzung aus der vorliegenden Arbeit, könnten entsprechende Projekte mit dessen deduktiv abgeleiteten Framework für die Konstruktion bspw. von Ratinginstrumenten oder Codiermanualen starten. So wäre ein theoriegeleitetes Vorgehen mit empirischer Grundierung möglich, das in der Pilotierung nicht ausschließlich auf induktive Erkenntnisse von betont qualitativen Instrumenten angewiesen wäre.

Eine stärker formal angelegte Erweiterung der Studie läge bei den *Regeln* als Ausgangspunkt für die Modellnutzung. Diese sind als Ausdruck formallogischer Argumente rekonstruierbar (Tetens, 2010) und insofern mit zwei relevanten Anschlussmöglichkeiten ausgestattet. Erstens erlauben Argumentationsrekonstruktionen die Rückführung von natürlicher Sprache in binäre Wahr-Falsch-Tabellen (vgl. Gilbert & Justi, 2016; Mendonça & Justi, 2014; Wittgenstein, 2006). Sie bieten damit einen direkten Messzugang auch auf qualitativ erhobene Denkprozesse mit Hilfe etablierter IRT-Modelle. Zweitens sind sie verallgemeinerbar und unabhängig davon, ob diese Denkprozesse den korrekten Umgang mit Inhaltswissen wiedergeben (Salmon, 2003; Tugendhat & Wolf, 1986). Rincke und Markic (2018) stellen Sprache im naturwissenschaftlichen Unterricht u. a. als Argumentationen im Unterricht dar. Dafür verweisen sie auf von Aufschnaiter und Precht (2018), die das Argumentieren jedoch weniger so auffassen wie in der vorliegenden Arbeit (vgl. Abschnitt 2.1.1), sondern als Diskurs im Sinne eines Meinungs-

austauschs. In dieser Hinsicht wird an dieser Stelle Forschungsbedarf gesehen. Wenn das Forschungsinteresse stärker auf dem individuellen Umgang mit Modellen und weniger auf der Lernwirksamkeit für den Erwerb von Fachinhalten liegt, dann besitzt der hier vorgeschlagene Verknüpfungsvorschlag das Potential einer Erweiterung des Forschungsfeldes in der Schnittmenge von Erkenntnisgewinnung und Sprache im naturwissenschaftlichen Unterricht.

Literatur

- Ackerman, T. A., Gierl, M. J. & Walker, C. M. (2003). Using Multidimensional Item Response Theory to Evaluate Educational and Psychological Tests. *Educational Measurement: Issues and Practice*, 22(3), 37–51. <https://doi.org/10.1111/j.1745-3992.2003.tb00136.x>
- Aduriz-Bravo, A. (2013). A Semantic View of Scientific Models for Science Education. *Science & Education*, 22(7), 1593–1611. <https://doi.org/10.1007/s11191-011-9431-7>
- Aduriz-Bravo, A. & Revel Chion, A. (2017). Language, Discourse, Argumentation, and Science Education. In K. S. Taber & B. Akpan (Hrsg.), *Science education* (S. 157–166). Sense publishers.
- Aliseda, A. (2000). Abduction As Epistemic Change. In P. A. Flach & A. C. Kakas (Hrsg.), *Abduction and Induction* (S. 45–85). Springer Science+Business Media B.V.
- Amrhein, V., Trafimow, D. & Greenland, S. (2019). Inferential Statistics as Descriptive Statistics. *The American Statistician*, 73(1), 262–270.
- Andersen, C. & Garcia-Mila, M. (2017). Scientific Reasoning During Inquiry. In K. S. Taber & B. Akpan (Hrsg.), *Science education* (S. 105–117). Sense publishers.
- APA. (2013). *Publication Manual of the American Psychological Association*. APA.
- Arnold, K. (Hrsg.). (2018). *Fokus Chemie 9/10*. Cornelsen.
- Arnold, K. & Dietrich, V. (Hrsg.). (2010). *Chemie plus*. Cornelsen.
- Bandalos, D. L. (2018). *Measurement Theory and Applications for the Social Sciences*. The Guilford Press.
- Bara, B. G. & Bucciarelli, M. (2000). Deduction and Induction. *Mind & Society*, 1(1), 95–107.
- Barke, H.-D., Harsch, G., Marohn, A. & Krees, S. (2015). *Chemididaktik kompakt (2.)*. Springer.
- Baumert, J., Lüdtke, O., Trautwein, U. & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition. *Educational Research Review*, 4(3), 165–176. <https://doi.org/10.1016/j.edurev.2009.04.002>
- Beckermann, A. (2001). Zur Inkohärenz und Irrelevanz des Wissensbegriffs. *Zeitschrift für Philosophische Forschung*, 55, 571–593.
- Behling, F., Förtsch, C. & Neuhaus, B. J. (2019). Sprachsensibler Biologieunterricht – Förderung professioneller Handlungskompetenz und professioneller Wahrnehmung durch videogestützte live-Unterrichtsbeobachtung. *Zeitschrift für Didaktik der Naturwissenschaften*, 25(1), 307–316. <https://doi.org/10.1007/s40573-019-00103-9>

- Bernholt, S., Neumann, K. & Parchmann, I. (2018). Learning Progressions. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 209–225). Springer.
- Bock, R. D. & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters. *Psychometrika*, *46*(4), 443–459. <https://doi.org/10.1007/BF02293801>
- Bodner, G. M. & Briggs, M. W. (2005). A Model of Molecular Visualization. In J. K. Gilbert (Hrsg.), *Visualization in science education* (S. 61–72). Springer.
- Bohrmann-Linde, C., Kröger, S. & Sier, I. (Hrsg.). (2018). *Chemie 1*. C. C. Buchner.
- Bolsinova, M. & Maris, G. (2016). Can IRT Solve the Missing Data Problem in Test Equating? *Frontiers in Psychology*, *6*, 1–13. <https://doi.org/10.3389/fpsyg.2015.01956>
- Bond, T. & Fox, C. M. (2015). *Applying the Rasch Model* (3rd ed.). Taylor and Francis.
- Bonifay, W. E., Reise, S. P., Scheines, R. & Meijer, R. R. (2015). When Are Multidimensional Data Unidimensional Enough for Structural Equation Modeling? *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(4), 504–516. <https://doi.org/10.1080/10705511.2014.938596>
- Boone, W. J. & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*(2), 253–269. <https://doi.org/10.1002/sce.20106>
- Bortz, J. & Döring, N. (Hrsg.). (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. vollständig überarbeitete, aktualisierte und erweiterte Auflage). Springer-Verlag.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Springer Berlin Heidelberg.
- Brady, C., Holbert, N., Soyulu, F., Novak, M. & Wilensky, U. (2015). Sandboxes for Model-Based Inquiry. *Journal of Science Education and Technology*, *24*(2-3), 265–286. <https://doi.org/10.1007/s10956-014-9506-8>
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3., aktualisierte und erw. Aufl.). Pearson Studium.
- Bühner, M., Kröner, S. & Ziegler, M. (2008). Working memory, visual–spatial-intelligence and their relationship to problem-solving. *Intelligence*, *36*(6), 672–680. <https://doi.org/10.1016/j.intell.2008.03.008>
- Burnham, K. P. & Anderson, D. R. (2004). Multimodel Inference. *Sociological Methods & Research*, *33*(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Byrne, R. M. J. & Tasso, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory & Cognition*, *27*(4), 726–740.
- Camilli, G. & Fox, J. (2015). An Aggregate IRT Procedure for Exploratory Factor Analysis. *Journal of Educational and Behavioral Statistics*, *40*(4), 377–401. <https://doi.org/10.3102/1076998615589185>
- Carlsen, W. S. (2007). Language and Science Learning. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of Research on Science Education* (S. 57–74). Lawrence Erlbaum Associates.
- Carstensen, C. H., Frey, A., Walter, O. & Knoll, S. (2007). Technische Grundlagen des dritten internationalen Vergleichs. In M. Prenzel, C. Artelt, J. Baumert, W.

- Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006* (S. 367–390). Waxmann.
- Caspari, I., Kranz, D. & Graulich, N. (2018). Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chem. Educ. Res. Pract.*, *19*(19), 1117–1141.
- Caspari, I., Weinrich, M. L., Sevian, H. & Graulich, N. (2018). This mechanistic step is productive: *Chem. Educ. Res. Pract.*, *19*(19), 42–59.
- Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6), 1–29.
- Chalmers, R. P. (2015). Extended Mixed-Effects Item Response Models With the MH-RM Algorithm. *Journal of Educational Measurement*, *52*(2), 200–222.
- Chalmers, R. P. & Ng, V. (2017). Plausible-Value Imputation Statistics for Detecting Item Misfit. *Applied Psychological Measurement*, *41*(5), 372–387. <https://doi.org/10.1177/0146621617692079>
- Chalmers, R. P., Pek, J. & Liu, Y. (2017). Profile-likelihood Confidence Intervals in Item Response Theory Models. *Multivariate behavioral research*, *52*(5), 533–550. <https://doi.org/10.1080/00273171.2017.1329082>
- Cheng, M.-F. & Lin, J.-L. (2015). Investigating the Relationship between Students' Views of Scientific Models and Their Development of Models. *International Journal of science education*, *37*(15), 2453–2475. <https://doi.org/10.1080/09500693.2015.1082671>
- Chon, K. H., Lee, W.-C. & Dunbar, S. B. (2010). A Comparison of Item Fit Statistics for Mixed IRT Models. *Journal of Educational Measurement*, *47*(3), 318–338.
- Cole, D. A. & Preacher, K. J. (2014). Manifest variable path analysis. *Psychological methods*, *19*(2), 300–315. <https://doi.org/10.1037/a0033805>
- Cooper, M. M. & Sandi-Urena, S. (2009). Design and Validation of an Instrument To Assess Metacognitive Skillfulness in Chemistry Problem Solving. *Journal of Chemical Education*, *86*(2), 240–245.
- Cooper, M. M., Sandi-Urena, S. & Stevens, R. (2008). Reliable multi method assessment of metacognition use in chemistry problem solving. *Chem. Educ. Res. Pract.*, *9*(1), 18–24.
- Crişan, D. R., Tendeiro, J. N. & Meijer, R. R. (2017). Investigating the Practical Consequences of Model Misfit in Unidimensional IRT Models. *Applied Psychological Measurement*, *41*(6), 439–455. <https://doi.org/10.1177/0146621617695522>
- Danusso, L., Testa, I. & Vicentini, M. (2010). Improving Prospective Teachers' Knowledge about Scientific Models and Modelling. *International Journal of science education*, *32*(7), 871–905. <https://doi.org/10.1080/09500690902833221>
- Dass, K., Head, M. L. & Rushton, G. T. (2015). Building an Understanding of How Model-Based Inquiry Is Implemented in the High School Chemistry Classroom. *Journal of Chemical Education*, *92*(8), 1306–1314. <https://doi.org/10.1021/acs.jchemed.5b00191>
- Devitt, M. (2006). Scientific Realism. In P. Greenough & M. P. Lynch (Hrsg.), *Truth and Realism* (S. 100–124). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199288878.003.0006>

- Dinno, A. (2015). Nonparametric Pairwise Multiple Comparisons in Independent Groups using Dunn's Test. *The Stata Journal: Promoting communications on statistics and Stata*, 15(1), 292–300. <https://doi.org/10.1177/1536867X1501500117>
- Dori, Y. J. & Kaberman, Z. (2012). Assessing high school chemistry students' modeling sub-skills in a computerized molecular modeling learning environment. *Instructional Science*, 40(1), 69–91. <https://doi.org/10.1007/s11251-011-9172-7>
- Döring, N. & Bortz, J. (2016a). Datenanalyse. In J. Bortz & N. Döring (Hrsg.), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 597–784). Springer-Verlag.
- Döring, N. & Bortz, J. (2016b). Datenaufbereitung. In J. Bortz & N. Döring (Hrsg.), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 579–595). Springer-Verlag.
- Döring, N. & Bortz, J. (2016c). Datenerhebung. In J. Bortz & N. Döring (Hrsg.), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 321–577). Springer-Verlag.
- Döring, N. & Bortz, J. (2016d). Untersuchungsdesign. In J. Bortz & N. Döring (Hrsg.), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (S. 181–220). Springer-Verlag.
- Dunbar, K. N. & Klahr, D. (2012). Scientific Thinking and Reasoning. In K. J. Holyoak & R. G. Morrison (Hrsg.), *The Oxford Handbook of Thinking and Reasoning*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199734689.013.0035>
- Erduran, S. & Kaya, E. (2019). *Transforming Teacher Education Through the Epistemic Core of Chemistry*. Springer International Publishing.
- Ergazaki, M. & Zogza, V. (2005). From A Causal Question To Stating And Testing Hypotheses. In K. Boersma, M. Goedhart, O. de Jong & H. Eijkelhof (Hrsg.), *Research and the Quality of Science Education* (S. 407–417). Springer.
- Espinet, M., Izquierdo, M., Bonil, J. & Ramos De Robles, S. L. (2012). The Role of Language in Modeling the Natural World. In B. J. Fraser, K. G. Tobin & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 1385–1403). Springer.
- European Centre for the Development of Vocational Training. (2019). 2018 skills forecast. <https://www.cedefop.europa.eu/en/publications-and-resources/country-reports/germany-2018-skills-forecast>
- Field, A., Miles, J. & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications.
- Flach, P. A. (2000). On The Logic Of Hypothesis Generation. In P. A. Flach & A. C. Kakas (Hrsg.), *Abduction and Induction* (S. 89–106). Springer Science+Business Media B.V.
- Flach, P. A. & Kakas, A. C. (2000). Abductive And Inductive Reasoning. In P. A. Flach & A. C. Kakas (Hrsg.), *Abduction and Induction* (S. 1–27). Springer Science+Business Media B.V.
- Flores-Camacho, F., Gallegos-Cázares, L., Garritz, A. & García-Franco, A. (2007). Incommensurability and Multiple Models. *Science & Education*, 16(7-8), 775–800. <https://doi.org/10.1007/s11191-006-9049-3>

- Fox, J. (2010). *Bayesian Item Response Modeling*. Springer.
- Fox, J. & Weisberg, S. (2019). An R Companion to Applied Regression. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Frey, A., Carstensen, C. H., Walter, O., Rönnebeck, S. & Gomolka, J. (2008). Methodische Grundlagen des Ländervergleichs. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006 in Deutschland* (S. 375–397). Waxmann.
- Frey, A., Hartig, J. & Rupp, A. A. (2009). An NCME Instructional Module on Booklet Designs in Large-Scale Assessments of Student Achievement. *Educational Measurement: Issues and Practice*, 28(3), 39–53. <https://doi.org/10.1111/j.1745-3992.2009.00154.x>
- Fulmer, G. W., Chu, H.-E., Treagust, D. F. & Neumann, K. (2015). Is it harder to know or to reason? *Asia-Pacific Science Education*, 1(1), 586. <https://doi.org/10.1186/s41029-015-0005-x>
- Gettier, E. L. (1963). Is Justified True Belief Knowledge? *Analysis*, 23(6), 121–123. <https://doi.org/10.1093/analys/23.6.121>
- Giere, R. N. (2004). How Models Are Used to Represent Reality. *Philosophy of Science*, 71(5), 742–752. <https://doi.org/10.1086/425063>
- Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, 172(2), 269–281. <https://doi.org/10.1007/s11229-009-9506-z>
- Gilbert, J. K. & Justi, R. (Hrsg.). (2016). *Modelling-based Teaching in Science Education* (Bd. 9). Springer.
- Gilbert, J. K. & Treagust, D. (Hrsg.). (2009). *Multiple Representations in Chemical Education* (Bd. 4). Springer Netherlands. <https://doi.org/10.1007/978-1-4020-8872-8>
- Glas, C. A. W., Pimentel, J. L. & Lamers, S. M. A. (2015). Nonignorable data in IRT models. *Psychological Test and Assessment Modeling*, 57(4), 523–541.
- Goodwin, R. M. (1967). A Growth Cycle. In C. H. Feinstein (Hrsg.), *Socialism, Capitalism and Economic Growth* (S. 54–58). Cambridge Univ. Press.
- Gouvea, J. & Passmore, C. (2017). ‘Models of’ versus ‘Models for’. *Science & Education*, 26(1), 49–63. <https://doi.org/10.1007/s11191-017-9884-4>
- Grosslight, L., Unger, C., Jay, E. & Smith, C. L. (1991). Understanding Models and their Use in Science. *Journal of Research in Science Teaching*, 28(9), 799–822.
- Grünkorn, J. (2014). *Modellkompetenz im Biologieunterricht* (Dissertation). Freie Universität Berlin. Berlin, Fachbereich Biologie, Chemie, Pharmazie.
- Gut-Glanzmann, C. & Mayer, J. (2018). Experimentelle Kompetenz. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 121–140). Springer.
- Han, Z. & Johnson, M. S. (2019). Global- and Item-Level Model Fit Indices. In M. von Davier & Y.-S. Lee (Hrsg.), *Handbook of Diagnostic Classification Models* (S. 265–285). Springer International Publishing.
- Harlen, W. (Hrsg.). (2015). *Working with Big Ideas of Science Education*. IAP for Science.

- Harlen, W. & Bell, D. (Hrsg.). (2010). *Principles and big ideas of science education*. Association for Science Education.
- Hartig, J. & Buchholz, J. (2012). A multilevel item response model for item position effects and individual persistence. *Psychological Test and Assessment Modeling*, *54*(4), 418–431.
- Hartig, J. & Frey, A. (2013). Sind Modelle der Item-Response-Theorie (IRT) das „Mittel der Wahl“ für die Modellierung von Kompetenzen? *Zeitschrift für Erziehungswissenschaft*, *16*(S1), 47–51. <https://doi.org/10.1007/s11618-013-0386-0>
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Springer Berlin Heidelberg.
- Hartig, J. & Höhler, J. (2008). Representation of Competencies in Multidimensional IRT Models with Within-Item and Between-Item Multidimensionality. *Zeitschrift für Psychologie / Journal of Psychology*, *216*(2), 89–101. <https://doi.org/10.1027/0044-3409.216.2.89>
- Hartig, J. & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, *35*(2-3), 57–63. <https://doi.org/10.1016/j.stueduc.2009.10.002>
- Hartig, J. & Höhler, J. (2010). Projekt MIRT. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung* (S. 189–198). Beltz.
- Härtig, H., Fraser, N., Bernholt, S. & Retelsdorf, J. (2019). Kann man Sachtexte vereinfachen? *Zeitschrift für Didaktik der Naturwissenschaften*, *25*(1), 273–287. <https://doi.org/10.1007/s40573-019-00105-7>
- Hartmann, S., Upmeyer zu Belzen, A., Krüger, D. & Pant, H. A. (2015). Scientific Reasoning in Higher Education. *Zeitschrift für Psychologie*, *223*(1), 47–53. <https://doi.org/10.1027/2151-2604/a000199>
- Harvey, R. J. (2016). Improving Measurement via Item Response Theory. *The Counseling Psychologist*, *44*(2), 195–204. <https://doi.org/10.1177/0011000015615427>
- Hayes, A. & Couch, S. (2020). broom: Convert Statistical Objects into Tidy Tibbles. <https://CRAN.R-project.org/package=broom>
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T. & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS biology*, *13*(3), e1002106. <https://doi.org/10.1371/journal.pbio.1002106>
- Heering, P. & Kremer, K. (2018). Nature of Science. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 105–119). Springer.
- Heller, K. A. & Perleth, C. (2000). Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (M. Hasselhorn, H. Marx & W. Schneider, Hrsg.).
- Henson, J. M., Reise, S. P. & Kim, K. H. (2007). Detecting Mixtures From Structural Model Differences Using Latent Variable Mixture Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(2), 202–226. <https://doi.org/10.1080/10705510709336744>
- Herrmann, S., Finke, F., Lülff, M., Kichakova, O., Puetzfeld, D., Knickmann, D., List, M., Rievers, B., Giorgi, G., Günther, C., Dittus, H., Prieto-Cerdeira, R., Dilssner, F., Gonzalez, F., Schönemann, E., Ventura-Traveset, J. & Lämmerzahl, C. (2018).

- Test of the Gravitational Redshift with Galileo Satellites in an Eccentric Orbit. *Physical review letters*, 121(23), 231102. <https://doi.org/10.1103/PhysRevLett.121.231102>
- Hui, E. G. M. (2019). *Learn R for Applied Statistics*. Apress.
- Jang, E. E. & Roussos, L. (2007). An Investigation into the Dimensionality of TOEFL Using Conditional Covariance-Based Nonparametric Approach. *Journal of Educational Measurement*, 44(1), 1–21. <https://doi.org/10.1111/j.1745-3984.2007.00024.x>
- Janiak, C. & Riedel, E. (2015). *Anorganische Chemie*. De Gruyter.
- Johnson-Laird, P. N. (1999). Deductive Reasoning. *Annu. Rev. Psychol.*, 50, 109–135.
- Johnson-Laird, P. N. (2010). Mental models and human reasoning. *PNAS*, 107(43), 18243–18250.
- Johnson-Laird, P. N., Byrne, R. M. J. & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, 99(3), 418–439.
- Johnstone, A. H. (2010). You Can't Get There from Here. *Journal of Chemical Education*, 87(1), 22–29.
- Jonkisz, E., Moosbrugger, H. & Brandt, H. (2012). Planung und Entwicklung von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 27–74). Springer Berlin Heidelberg.
- Kassambara, A. (2020). ggpubr: ggplot2 Based Publication Ready Plots. <https://CRAN.R-project.org/package=ggpubr>
- Kauertz, A. (2014). Entwicklung eines Rasch-skalierten Leistungstests. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 341–353). Springer Berlin Heidelberg.
- Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E. & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 135–153.
- Khemlani, S. S., Mackiewicz, R., Bucciarelli, M. & Johnson-Laird, P. N. (2013). Kinematic mental simulations in abduction and deduction. *PNAS*, 110(42), 16766–16771.
- Kind, P. & Osborne, J. (2017). Styles of Scientific Reasoning. *Science Education*, 101(1), 8–31. <https://doi.org/10.1002/sce.21251>
- Kirch, S. A. (2012). Understanding Scientific Uncertainty as a Teaching and Learning Goal. In B. J. Fraser, K. G. Tobin & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 851–864). Springer.
- Klein, U. (2001). Berzelian Formulas as Paper Tools in Early Nineteenth-Century Chemistry. *Foundations of Chemistry*, 3(1), 7–32. <https://doi.org/10.1023/A:1011460318779>
- Klein, U. (2003). *Experiments, Models, Paper Tools*. Stanford University Press.
- Klüver, J. (2015). Modellbildung und wissenschaftliche Praxis. *Erwägen Wissen Ethik*, 26(3), 391–394.
- Knuuttila, T. (2005). *Models as Epistemic Artefacts* (Bd. 8). Edita Prima Oy.
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42, 262–271.

- Knuuttila, T. (2017). Imagination extended and embedded. *Synthese*, 99(3), 56. <https://doi.org/10.1007/s11229-017-1545-2>
- Knuuttila, T. & Boon, M. (2011). How to models give us knowledge? The case of Carnot's ideal heat engine. *European Journal for Philosophy of Science*, (1), 309–334. <https://doi.org/10.1007/s13194-011-0029-3>
- Knuuttila, T. & Loettgers, A. (2014). Magnets, Spins, and Neurons: The Dissemination of Model Templates Across Disciplines. *The Monist*, 97(3), 280–300.
- Krell, M. (2018). Schwierigkeitserzeugende Aufgabenmerkmale bei Multiple-Choice-Aufgaben zur Experimentierkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1), 1–15. <https://doi.org/10.1007/s40573-017-0069-0>
- Krell, M., Reinisch, B. & Krüger, D. (2015). Analyzing Students' Understanding of Models and Modeling Referring to the Disciplines Biology, Chemistry, and Physics. *Research in Science Education*, 45(3), 367–393. <https://doi.org/10.1007/s11165-014-9427-9>
- Kress, G. R. & van Leeuwen, T. (2001). *Multimodal Discourse*. Bloomsbury Academic.
- Krüger, D., Kauertz, A. & Upmeyer zu Belzen, A. (2018). Modelle und das Modellierung in den Naturwissenschaften. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 141–157). Springer.
- Krüger, D., Parchmann, I. & Schecker, H. (Hrsg.). (2014). *Methoden in der naturwissenschaftsdidaktischen Forschung*. Springer Berlin Heidelberg.
- Landesinstitut für Schule und Medien Berlin-Brandenburg (Hrsg.). (2016). Rahmenlehrplan Online. Verfügbar 25. Juni 2016 unter <http://bildungsserver.berlin-brandenburg.de/rlp-online/startseite/>
- Landesinstitut für Schule und Medien Berlin-Brandenburg (Hrsg.). (2017). Rahmenlehrplan Online. Verfügbar 18. September 2019 unter <http://bildungsserver.berlin-brandenburg.de/rlp-online/startseite/>
- Laszlo, P. (2013). Towards Teaching Chemistry as a Language. *Science & Education*, 22(7), 1669–1706. <https://doi.org/10.1007/s11191-011-9408-6>
- Laukaiyte, I. & Wiberg, M. (2017). Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics - Theory and Methods*, 46(22), 11341–11357. <https://doi.org/10.1080/03610926.2016.1267764>
- Lederman, N. G. & Abd-El-Khalick. (2002). Avoiding De-Natured Science. In W. F. McComas (Hrsg.), *The Nature of Science in Science Education* (S. 83–126). Kluwer Academic Publishers.
- Lee, S., Sriutaisuk, S. & Kim, H. (2020). Using the Tidyverse Package in R for Simulation Studies in SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(3), 468–482. <https://doi.org/10.1080/10705511.2019.1644515>
- Lehrer, R. & Schauble, L. (2003). Origins and Evolution of Model-Based Reasoning in Mathematics and Science. In R. A. Lesh & H. M. Doerr (Hrsg.), *Beyond constructivism* (S. 59–70). Lawrence Erlbaum Associates.
- Lesh, R. A. (2000). Beyond Constructivism: Identifying Mathematical Abilities that are Most Needed for Success Beyond School in an Age of Information. *Mathematics Education Research Journal*, 12(3), 177–195.

- Lesh, R. A., Doerr, H. M., Carmona, G. & Hjalmarson, M. (2003). Beyond Constructivism. *Mathematical Thinking and Learning*, 5(2-3), 211–233.
- Lesh, R. A., Hoover, M., Hole, B., Kelly, A. & Post, T. (2000). Principles for Developing Thought-Revealing Activities for Students and Teachers. In A. Kelly & R. A. Lesh (Hrsg.), *Handbook of Research Design in Mathematics and Science Education* (S. 591–646). Lawrence Erlbaum Associates.
- Lesh, R. A. & Zawojewski, J. (2007). Problem Solving and Modeling. In F. K. Lester (Hrsg.), *The Second Handbook of Research on Mathematics Teaching and Learning* (S. 763–804). Information Age Publishing.
- Levy, R. (2011). Bayesian Data-Model Fit Assessment for Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(4), 663–685. <https://doi.org/10.1080/10705511.2011.607723>
- Lorah, J. & Womack, A. (2019). Value of sample size for computation of the Bayesian information criterion (BIC) in multilevel modeling. *Behavior Research Methods*, 51(1), 440–450. <https://doi.org/10.3758/s13428-018-1188-3>
- Louca, L. T. & Zacharia, Z. C. (2012). Modeling-based learning in science education. *Educational Review*, 64(4), 471–492. <https://doi.org/10.1080/00131911.2011.628748>
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58(2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>
- Luxford, C. J. & Bretz, S. L. (2013). Moving beyond definitions. *Chem. Educ. Res. Pract.*, 14(2), 214–222. <https://doi.org/10.1039/c3rp20154f>
- Mahr, B. (2008). Ein Modell des Modellseins. In U. Dirks & E. Knobloch (Hrsg.), *Modelle* (S. 187–218). Lang.
- Mahr, B. (2011). On the Epistemology of Models. In G. Abel & J. Conant (Hrsg.), *Rethinking Epistemology* (S. 301–352). De Gruyter.
- Mahr, B. (2015). Modelle und ihre Befragbarkeit. *Erwägen Wissen Ethik*, 26(3), 329–342.
- Mangiafico, S. (2020). rcompanion: Functions to Support Extension Education Program Evaluation. <https://CRAN.R-project.org/package=rcompanion>
- Matthews, M. R. (2007). Models in science and in science education. *Science & Education*, 16(7-8), 647–652. <https://doi.org/10.1007/s11191-007-9089-3>
- Maydeu-Olivares, A. & Joe, H. (2005). Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2 n Contingency Tables. *Journal of the American Statistical Association*, 100(471), 1009–1020. <https://doi.org/10.1198/016214504000002069>
- Maydeu-Olivares, A. & Joe, H. (2006). Limited Information Goodness-of-fit Testing in Multidimensional Contingency Tables. *Psychometrika*, 71(4), 713–732. <https://doi.org/10.1007/s11336-005-1295-9>
- Mayer, J. & Wellnitz, N. (2014). Die Entwicklung von Kompetenzstrukturmodellen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 19–29). Springer Berlin Heidelberg.

- McCarthy, C. L. (2018). Cultural Studies of Science Education. In M. R. Matthews (Hrsg.), *History, Philosophy and Science Teaching* (S. 99–136). Springer International Publishing.
- McElreath, R. (2020). *Statistical Rethinking* (Second edition). CRC Press.
- Meinel, C. (2008). Kugeln und Stäbchen: Vom kulturellen Ursprung chemischer Molekülmodelle. In U. Dirks & E. Knobloch (Hrsg.), *Modelle* (S. 221–234). Lang.
- Mendonça, P. C. C. & Justi, R. (2014). An instrument for analyzing arguments produced in modeling-based chemistry lessons. *Journal of Research in Science Teaching*, *51*(2), 192–218. <https://doi.org/10.1002/tea.21133>
- Mizrahi, M. (2011). Why the ultimate argument for scientific realism ultimately fails. *Studies in History and Philosophy of Science*, *43*, 132–138. <https://doi.org/10.1016/j.shpsa.2011.11.001>
- Moosbrugger, H. (2012). Item-Response-Theory (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 227–274). Springer Berlin Heidelberg.
- Moosbrugger, H. & Kelava, A. (2012a). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 7–26). Springer Berlin Heidelberg.
- Moosbrugger, H. & Kelava, A. (Hrsg.). (2012b). *Testtheorie und Fragebogenkonstruktion*. Springer Berlin Heidelberg.
- Moreira, P., Marzabal, A. & Talanquer, V. (2019). Using a mechanistic framework to characterise chemistry students' reasoning in written explanations. *Chem. Educ. Res. Pract.*, *20*(1), 120–131. <https://doi.org/10.1039/c8rp00159f>
- National Research Council (Hrsg.). (2012). *A Framework for K-12 Science Education*. The National Academies Press.
- Nehring, A. (2014). *Wissenschaftliche Denk- und Arbeitsweisen im Fach Chemie* (Bd. 177). Logos Verl.
- Nehring, A., Nowak, K. H., Upmeier zu Belzen, A. & Tiemann, R. (2015). Predicting Students' Skills in the Context of Scientific Inquiry with Cognitive, Motivational, and Sociodemographic Variables. *International Journal of science education*, *37*(9), 1343–1363. <https://doi.org/10.1080/09500693.2015.1035358>
- Nehring, A. & Schwichow, M. (2020). Was ist Wissen, was ist Können? *Zeitschrift für Didaktik der Naturwissenschaften*, *36*(16), 2719. <https://doi.org/10.1007/s40573-020-00113-y>
- Nehring, A., Stiller, J., Nowak, K. H., Upmeier zu Belzen, A. & Tiemann, R. (2016). Naturwissenschaftliche Denk- und Arbeitsweisen im Chemieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, (22), 77–96.
- Neumann, I. (2011). *Beyond Physics Content Knowledge* (Bd. 117). Logos Verl.
- Neumann, K. (2014). Rasch-Analyse naturwissenschaftsbezogener Leistungstests. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschafts-didaktischen Forschung* (S. 355–369). Springer Berlin Heidelberg.
- Nicolaou, C. T. & Constantinou, C. P. (2014). Assessment of the modeling competence. *Educational Research Review*, *13*, 52–73.

- Nitz, S. & Fechner, S. (2018). Mentale Modelle. In D. Krüger, I. Parchmann & H. Schekker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 69–86). Springer.
- OECD. (2019a). PISA 2018 Results (Volume I). <https://doi.org/10.1787/5f07c754-en>
- OECD. (2019b). PISA 2018 Results (Volume II). <https://doi.org/10.1787/b5fd1b8f-en>
- Ogle, D. H., Wheeler, P. & Dinno, A. (2020). FSA: Fisheries Stock Analysis. <https://github.com/droglenc/FSA>
- Oliva, J. M., del Mar Aragón, M. & Cuesta, J. (2015). The Competence of Modelling in Learning Chemical Change. *International Journal of Science and Mathematics Education*, 13(4), 751–791.
- O'Reilly, T. & McNamara, D. S. (2007). The Impact of Science Knowledge, Reading Skill, and Reading Strategy Knowledge on More Traditional “High-Stakes” Measures of High School Students’ Science Achievement. *American Educational Research Journal*, 44(1), 161–196. <https://doi.org/10.3102/0002831206298171>
- Orlando, M. & Thissen, D. (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. *Applied Psychological Measurement*, 24(1), 50–64.
- Orlando, M. & Thissen, D. (2003). Further Investigation of the Performance of S - X2. *Applied Psychological Measurement*, 27(4), 289–298. <https://doi.org/10.1177/0146621603027004004>
- Osborne, J. (2012). The Role of Argument: Learning How to Learn in School Science. In B. J. Fraser, K. G. Tobin & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 933–949). Springer.
- Osborne, J. & Patterson, A. (2011). Scientific argument and explanation. *Science Education*, 95(4), 627–638. <https://doi.org/10.1002/sce.20438>
- Paavola, S. (2004). Abduction As A Logic And Methodology Of Discovery. *Foundation of Science*, 9, 267–283.
- Paek, I. & Cole, K. (2020). *Using R for item response theory model applications*. Routledge.
- Paek, I., Cui, M., Öztürk G., N. & Yang, Y. (2018). Estimation of an IRT Model by Mplus for Dichotomously Scored Responses Under Different Estimation Methods. *Educational and psychological measurement*, 78(4), 569–588. <https://doi.org/10.1177/0013164417715738>
- Parchmann, I. (2010). Kompetenzmodellierung in den Naturwissenschaften. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung* (S. 135–142). Beltz.
- Passmore, C., Gouvea, J. & Giere, R. N. (2014). Models in Science and in Learning Science. In M. R. Matthews (Hrsg.), *International Handbook of Research in History, Philosophy and Science Teaching* (S. 1171–1202). Springer Netherlands.
- Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C. & Tsourlidaki, E. (2015). Phases of inquiry-based learning. *Educational Research Review*, 14, 47–61.
- Peng, R. D. & Matsui, E. (2016). *The Art of Data Science*.

- Penner, D. E., Giles, N. D., Lehrer, R. & Schauble, L. (1997). Building Functional Models: Designing an Elbow. *Journal of Research in Science Teaching*, 34(2), 125–143.
- Pohl, S., Gräfe, L. & Rose, N. (2013). Dealing With Omitted and Not-Reached Items in Competence Tests. *Educational and psychological measurement*, 74(3), 423–452. <https://doi.org/10.1177/0013164413504926>
- Popper, K. (1935). *Logik der Forschung*. Springer Vienna.
- Portides, D. P. (2007). The Relation between Idealisation and Approximation in Scientific Model Construction. *Science & Education*, 16(7-8), 699–724. <https://doi.org/10.1007/s11191-006-9001-6>
- R Core Team. (2020). R: A language and environment for statistical computing. <https://www.R-project.org>
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. Springer.
- Reed, D. K., Petscher, Y. & Truckenmiller, A. J. (2017). The Contribution of General Reading Ability to Science Achievement. *Reading Research Quarterly*, 52(2), 253–266. <https://doi.org/10.1002/rrq.158>
- Revelle, W. (2019). psych: Procedures for Psychological, Psychometric, and Personality Research. <https://CRAN.R-project.org/package=psych>
- Riemeier, T. (2007). Moderater Konstruktivismus. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung* (S. 69–79). Springer-Verlag Berlin Heidelberg.
- Riese, J. & Reinhold, P. (2014). Entwicklung eines Leistungstests für fachdidaktisches Wissen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 257–267). Springer Berlin Heidelberg.
- Rinke, K. & Markic, S. (2018). Sprache und das Lernen von Naturwissenschaften. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 31–48). Springer.
- Robitzsch, A., Kiefer, T. & Wu, M. (2020). TAM: Test Analysis Modules. <https://CRAN.R-project.org/package=TAM>
- Robitzsch, A., Lüdtke, O., Köller, O., Kröhne, U., Goldhammer, F. & Heine, J.-H. (2017). Herausforderungen bei der Schätzung von Trends in Schulleistungsstudien. *Diagnostica*, 63(2), 148–165. <https://doi.org/10.1026/0012-1924/a000177>
- Rönnebeck, S., Bernholt, S. & Ropohl, M. (2016). Searching for a common ground. *Studies in Science Education*, 52(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>
- Ropohl, M., Walpuski, M. & Sumfleth, E. (2015). Welches Aufgabenformat ist das richtige? *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 1–15. <https://doi.org/10.1007/s40573-014-0020-6>
- Rose, N., Nagy, G., Nagengast, B., Frey, A. & Becker, M. (2019). Modeling Multiple Item Context Effects With Generalized Linear Mixed Models. *Frontiers in Psychology*, 10, 248. <https://doi.org/10.3389/fpsyg.2019.00248>
- Rost, J. (2006). Item-Response-Theorie. In F. Petermann, M. Eid & J. Bengel (Hrsg.), *Handbuch der psychologischen Diagnostik* (S. 261–274). Hogrefe.

- Rost, M. & Tiemann, R. (2017). Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht. In J. Stiller & C. Laschke (Hrsg.), *Berlin-Brandenburger Beiträge zur Bildungsforschung* (S. 283–303). Peter Lang.
- Rost, M. & Tiemann, R. (2020). Empirische Struktur von Modellnutzung im Chemieunterricht. In S. Habig (Hrsg.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen* (S. 532–535).
- Roth, K. (2005). Eine unendliche chemische Geschichte. *Chemie in unserer Zeit*, 39, 212–217.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.
- Saint-Mont, U. (2011). *Statistik im Forschungsprozess*. Physica-Verlag HD.
- Salmon, W. C. (2003). *Logik* (Bd. 7996). Reclam.
- Sauer, S. (2019). *Moderne Datenanalyse mit R*. Springer Fachmedien Wiesbaden.
- Scantlebury, K. (2012). Still Part of the Conversation. In B. J. Fraser, K. G. Tobin & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 499–512). Springer.
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45–66.
- Schecker, H., Parchmann, I. & Krüger, D. (2014). Formate und Methoden naturwissenschaftsdidaktischer Forschung. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 1–15). Springer Berlin Heidelberg.
- Scherer, R. (2011). Covariates of complex problem solving competency in chemistry. *Science Education Review Letters*, 1–6.
- Scherer, R. (2014). Komplexes Problemlösen im Fach Chemie. *Zeitschrift für Pädagogische Psychologie*, 28(4), 181–192. <https://doi.org/10.1024/1010-0652/a000136>
- Scherer, R. & Beckmann, J. F. (2014). The acquisition of problem solving competence. *Large-scale Assessments in Education*, 2(1), 397. <https://doi.org/10.1186/s40536-014-0010-7>
- Scherer, R. & Tiemann, R. (2014). Measuring Students' Progressions in Scientific Problem Solving. *Procedia - Social and Behavioral Sciences*, 112, 87–96. <https://doi.org/10.1016/j.sbspro.2014.01.1142>
- Schermelleh-Engel, K., Kelava, A. & Moosbrugger, H. (2006). Gütekriterien. In F. Petermann, M. Eid & J. Bengel (Hrsg.), *Handbuch der psychologischen Diagnostik* (S. 420–433). Hogrefe.
- Schermelleh-Engel, K. & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 119–141). Springer Berlin Heidelberg.
- Schipolowski, S., Wittig, J., Mahler, N. & Stanat, P. (2019). Geschlechtsbezogene Disparitäten. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Hentschel (Hrsg.), *IQB-Bildungstrend 2018* (S. 237–263). Waxmann.
- Schmiemann, P. & Lücken, M. (2014). Validität. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 107–118). Springer Berlin Heidelberg.

- Schneider, W., Schlagmüller, M. & Ennemoser, M. (2007). Lesegeschwindigkeits- und verständnistest für die Klassen 6-12.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B. & Krajcik, J. (2009). Developing a learning progression for scientific modeling. *Journal of Research in Science Teaching*, 46(6), 632–654. <https://doi.org/10.1002/tea.20311>
- Schwarz, C. V. & White, B. Y. (2005). Metamodeling Knowledge. *Cognition and Instruction*, 23(2), 165–205. https://doi.org/10.1207/s1532690xci2302_1
- Schwichow, M. & Nehring, A. (2018). Variablenkontrolle beim Experimentieren in Biologie, Chemie und Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1), 217–233. <https://doi.org/10.1007/s40573-018-0085-8>
- Seo, D. G., Choi, Y. & Huh, S. (2017). Usefulness of the DETECT program for assessing the internal structure of dimensionality in simulated data and results of the Korean nursing licensing examination. *Journal of educational evaluation for health professions*, 14, 32. <https://doi.org/10.3352/jeehp.2017.14.32>
- Shavelson, R. J. (2013). On an Approach to Testing and Modeling Competence. *Educational Psychologist*, 48(2), 73–86. <https://doi.org/10.1080/00461520.2013.779483>
- Simms, N. K. & Richland, L. E. (2019). Generating Relations Elicits a Relational Mindset in Children. *Cognitive Science*, 43(10). <https://doi.org/10.1111/cogs.12795>
- Spieß, M. (2010). Der Umgang mit fehlenden Werten. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 117–142). VS Verlag für Sozialwissenschaften.
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Springer-Verlag.
- Stachowiak, H. (1983). Erkenntnisstufen zum Systematischen Neopragmatismus und zur Allgemeinen Modelltheorie. In H. Stachowiak (Hrsg.), *Modelle* (S. 87–146). Fink.
- Stephens, A. L. & Clement, J. J. (2012). The Role of Thought Experiments in Science and Science Learning. In B. J. Fraser, K. G. Tobin & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 157–175). Springer.
- Stewart, J. (2012). Does IRT Provide more sensitive measures of latent traits in statistical tests? *Shiken Research Bulletin*, 16(1), 15–22.
- Stieff, M., Scopelitis, S., Lira, M. E. & Desutter, D. (2016). Improving Representational Competence with Concrete Models. *Science Education*, 100(2), 344–363.
- Stone, C. A. & Zhang, B. (2003). Assessing Goodness of Fit of Item Response Theory Models. *Journal of Educational Measurement*, 40(4), 331–352.
- Streller, S., Bolte, C., Dietz, D. & La Noto Diega, R. (2019). *Chemiedidaktik an Fallbeispielen*. Springer.
- Suhr, D. (2017). *Die Alchemisten*. Springer.
- Sumfleth, E. & Nakoinz, S. (2019). Chemie verstehen – beobachtbare makroskopische Phänomene auf submikroskopischer Ebene modellbasiert interpretieren. *Zeitschrift für Didaktik der Naturwissenschaften*, 25(1), 231–243. <https://doi.org/10.1007/s40573-019-00101-x>
- Terzer, E., Hartig, J. & Upmeyer zu Belzen, A. (2013). Systematische Konstruktion eines Tests zu Modellkompetenz im Biologieunterricht unter Berücksichtigung von Gütekriterien. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 51–76.

- Tetens, H. (2010). *Philosophisches Argumentieren* (Bd. 1607). Beck.
- Tiemann, R. & Körbs, C. (2014). Die Fragebogenmethode, ein Klassiker der empirischen didaktischen Forschung. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 283–295). Springer Berlin Heidelberg.
- Tugendhat, E. & Wolf, U. (1986). *Logisch-semantische Propädeutik*. Reclam.
- Upmeier zu Belzen, A. & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57.
- Upmeier zu Belzen, A., Krüger, D. & van Driel, J. (Hrsg.). (2019). *Towards a Competence-Based View on Models and Modeling in Science Education* (Bd. 12). Springer International Publishing. <https://doi.org/10.1007/978-3-030-30255-9>
- van Fraassen, B. C. (1980). *The Scientific Image*. Clarendon Press.
- Vock, M. & Holling, H. (2006). Intelligenzdiagnostik. In F. Petermann, M. Eid & J. Bengel (Hrsg.), *Handbuch der psychologischen Diagnostik* (S. 494–502). Hogrefe.
- von Davier, M. (2014). The Log-Linear Cognitive Diagnostic Model (LCDM) as a Special Case of the General Diagnostic Model (GDM). *ETS Research Report Series*, (2), 1–13.
- von Davier, M., Gonzalez, E. & Mislevy, R. J. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Hrsg.), *Issues and Methodologies in Large-Scale Assessments* (S. 9–36).
- von Davier, M. & Hastedt, D. (Hrsg.). (2009). *Issues and Methodologies in Large-Scale Assessments* (Bd. 2).
- von Aufschnaiter, C. & Precht, H. (2018). Argumentieren im naturwissenschaftlichen Unterricht. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 87–104). Springer.
- Vorholzer, A., von Aufschnaiter, C. & Kirschner, S. (2016). Entwicklung und Erprobung eines Tests zur Erfassung des Verständnisses experimenteller Denk- und Arbeitsweisen. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 25–41. <https://doi.org/10.1007/s40573-015-0039-3>
- Vosniadou, S. (2012). Reframing the Classical Approach to Conceptual Change. In B. J. Fraser, K. G. Tobin & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 119–130). Springer.
- Walpuski, M. & Ropohl, M. (2014). Statistische Verfahren für die Analyse des Einflusses von Aufgabenmerkmalen auf die Schwierigkeit. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 385–398). Springer Berlin Heidelberg.
- Warm, T. A. (1989). Weighted Likelihood Estimation of Ability in Item Response Theory. *Psychometrika*, 54(3), 427–450.
- Wasserstein, R. L. & Lazar, N. A. (2016). The ASA's Statement on p -Values. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- Weirich, S. (2015). *Kontexteffekte in Large-Scale Assessments* (Dissertation). Humboldt-Universität zu Berlin. Berlin.
- Weyer, J. (2018). *Geschichte der Chemie Band 1* (Bd. 1). Springer Berlin Heidelberg.

- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grole-
mund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E.,
Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani,
H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43),
1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H. & Grolemund, G. (n. d.). R for Data Science. <https://r4ds.had.co.nz/>
- Wittgenstein, L. (2006). *Werkausgabe Band 1* (20. Auflage, Bd. 501). Suhrkamp.
- Woitkowski, D. (2015). *Fachliches Wissen Physik in der Hochschulausbildung* (Bd. 185).
Logos Verlag Berlin GmbH.
- Woitkowski, D. & Riese, J. (2017). Kriterienorientierte Konstruktion eines Kompetenz-
niveaumodells im physikalischen Fachwissen. *Zeitschrift für Didaktik der Natur-
wissenschaften*, 23(1), 39–52. <https://doi.org/10.1007/s40573-016-0054-z>
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational
Evaluation*, 31(2-3), 114–128. <https://doi.org/10.1016/j.stueduc.2005.05.005>
- Wu, M. (2010). Measurement, Sampling, and Equating Errors in Large-Scale Assess-
ments. *Educational Measurement: Issues and Practice*, 29(4), 15–27. <https://doi.org/10.1111/j.1745-3992.2010.00190.x>
- Wu, M., Tam, H. P. & Jen, T.-H. (2016). *Educational Measurement for Applied Research-
ers*. Springer Singapore.
- Xie, Y. (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R.
- Yang, F.-Y. & Tsai, C.-C. (2012). Personal Epistemology and Science Learning. In B. J.
Fraser, K. G. Tobin & C. J. McRobbie (Hrsg.), *Second International Handbook
of Science Education* (S. 259–280). Springer.
- Zhang, B. & Stone, C. A. (2007). Evaluating Item Fit for Multidimensional Item Respon-
se Models. *Educational and psychological measurement*, 68(2), 181–196. <https://doi.org/10.1177/0013164407301547>
- Zhang, J. (2013). A procedure for dimensionality analyses of response data from various
test designs. *Psychometrika*, 78(1), 37–58. <https://doi.org/10.1007/S11336-012-9287-Z>
- Zhang, J., Lu, J., Chen, F. & Tao, J. (2019). Exploring the Correlation Between Multiple
Latent Variables and Covariates in Hierarchical Data Based on the Multilevel
Multidimensional IRT Model. *Frontiers in Psychology*, 10, 2387. <https://doi.org/10.3389/fpsyg.2019.02387>
- Ziepprecht, K., Schwanewedel, J., Heitmann, P., Jansen, M., Fischer, H. E., Kauertz,
A., Kobow, I., Mayer, J., Sumfleth, E. & Walpuski, M. (2017). Modellierung
naturwissenschaftlicher Kommunikationskompetenz. *Zeitschrift für Didaktik der
Naturwissenschaften*, 23(1), 113–125. <https://doi.org/10.1007/s40573-017-0061-8>
- Zoller, U. & Nahum, T. L. (2012). From Teaching to KNOW to Learning to THINK
in Science Education. In B. J. Fraser, K. G. Tobin & C. J. McRobbie (Hrsg.),
Second International Handbook of Science Education (S. 209–229). Springer.

A. Abkürzungsverzeichnis

1-/2-PL 1-/2-Parameter Logistisches (Modell)
AIC Akaike Information Criterion
AICc bias corrected Akaike Information Criterion
AR Auswertung/Reflexion
APA American Psychological Association
BIBD balanced incomplete test design
BIC Bayesian Information Criterion
COV Kovariable(n)
DS Data Science
EAP-PV Expected-A-Posteriori Plausible-Value
FH Fragestellung/Hypothese
FW Fachwissenstest
HOCS higher-order cognitive skills
ICBD incomplete block design
ICC Item Characteristic Curve
IRT Item Response Theorie
KFT Kognitiver Fähigkeitstest
kIS kleinste Sinneinheiten
KTt klassische Testtheorie
LGVT Lesegeschwindigkeits- und Verständnistest
LOCS lower-order cognitive skills
logLik logarithmierte Likelihood
LR Likelihood Ratio
MAR Missing-At-Random
MC Multiple-Choice
MCAR Missing-Completely-At-Random
MDIFF multidimensionale Aufgabenschwierigkeit
MDISC multidimensionale Aufgabentrennschärfe
ML Maximum Likelihood
MNAR Missing-Not-At-Random
MNSQ Standardized Unweighted Mean Square
Mod Modellnutzungsfähigkeit
OECD Organisation for Economic Co-operation and Development
Ope Operationen
PD Planung/Durchführung
PISA Programme for International Student Assessment
PSE Periodensystem der Elemente
Rel Relationen
Reg Regeln
saBIC sample size adjusted Bayesian Information Criterion
SuS Schülerinnen und Schüler
WLE Weighted Maximum Likelihood Estimate
wMNSQ Standardized Weighted Mean Square

B. Aufgabenparameter

B.1. Schwierigkeitsparameter

Tabelle B.1.: Schwierigkeitsparameter aus den verschiedenen Skalierungen, inkl. multi-dim. Aufgabenschwierigkeiten aus der 4-dimensionalen Skalierung.

Item	M	ξ_{1d}	SE(ξ_{1d})	ξ_{3d}	SE(ξ_{3d})	ξ_{4d}	SE(ξ_{4d})	MDIFF _{4d}
BINFHpr	0.500	0.011	0.186	0.002	0.187	-0.012	0.184	-0.009
BINFHre	0.431	0.424	0.199	0.404	0.198	0.430	0.200	0.292
BINFHop	0.395	0.419	0.184	0.413	0.184	0.430	0.184	0.899
BINFHru	0.218	1.548	0.257	1.536	0.258	2.869	0.346	2.435
BINPDpr	0.654	-0.745	0.195	-0.694	0.196	-0.746	0.197	-0.717
BINPDre	0.359	0.714	0.196	1.859	0.314	1.503	0.294	0.360
BINPDop	0.626	-0.566	0.193	-0.554	0.192	-0.570	0.193	-0.445
BINPDru	0.229	1.331	0.240	1.611	0.261	1.507	0.252	1.277
BINARpr	0.299	0.865	0.195	0.865	0.195	0.859	0.194	0.782
BINARre	0.167	1.728	0.241	1.789	0.244	1.764	0.242	2.426
BINARop	0.519	-0.127	0.176	-0.455	0.295	-0.147	0.180	-0.292
BINARru	0.398	0.427	0.200	0.448	0.208	0.623	0.261	0.664
DESFHpr	0.299	1.541	0.243	1.432	0.237	1.405	0.235	0.744
DESFHre	0.541	-0.187	0.191	-0.183	0.187	-0.242	0.204	-0.157
DESFHop	0.476	0.240	0.207	0.222	0.206	0.158	0.221	0.319
DESFHru	0.254	1.149	0.206	1.130	0.205	1.287	0.214	5.733
DESPDpr	0.473	0.123	0.184	0.140	0.193	0.116	0.183	0.141
DESPDre	0.303	0.811	0.192	0.956	0.200	0.834	0.193	1.195
DESPDop	0.482	0.090	0.198	0.116	0.199	0.094	0.202	0.139
DESPDru	0.450	0.281	0.191	0.345	0.208	0.385	0.217	0.312
DESARpr	0.361	0.580	0.189	0.624	0.195	0.594	0.190	0.422
DESARre	0.273	1.149	0.216	1.125	0.214	1.108	0.214	0.642
DESARop	0.514	-0.120	0.207	-0.110	0.207	-0.260	0.244	-0.157
DESARru	0.546	-0.255	0.183	-0.273	0.186	-0.302	0.186	-0.385
DIPFHpr	0.326	0.871	0.212	0.890	0.209	0.900	0.210	0.492
DIPFHre	0.324	0.726	0.213	0.743	0.212	0.775	0.213	0.501
DIPFHop	0.414	0.443	0.209	0.464	0.204	0.433	0.203	0.252
DIPFHru	0.200	1.416	0.218	1.420	0.218	2.068	0.261	2.921
DIPPDpr	0.504	-0.076	0.185	-0.085	0.185	-0.071	0.185	-0.042

Anm.: M = deskr. Schw., MDIFF = multidim. Schw., Forts.: Folgeseite.

B. Aufgabenparameter

Tabelle B.1.: Schwierigkeitsparameter aus den verschiedenen Skalierungen, inkl. multi-dim. Aufgabenschwierigkeiten aus der 4-dimensionalen Skalierung (Fortsetzung).

Item	M	ξ_{1d}	SE(ξ_{1d})	ξ_{3d}	SE(ξ_{3d})	ξ_{4d}	SE(ξ_{4d})	MDIFF _{4d}
DIPPDre	0.274	0.983	0.218	0.988	0.220	1.005	0.219	1.107
DIPPDop	0.589	-0.412	0.188	-0.427	0.191	-1.492	0.337	-0.366
DIPPDru	0.306	0.869	0.198	0.854	0.197	1.097	0.222	1.145
DIPARpr	0.600	-0.548	0.218	-0.515	0.222	-0.530	0.211	-0.317
DIPARre	0.490	0.175	0.208	0.474	0.300	0.134	0.217	0.145
DIPARop	0.468	0.145	0.180	0.176	0.198	0.131	0.203	0.510
DIPARru	0.586	-0.313	0.184	-0.371	0.202	-0.851	0.353	-0.296
RKGFHpr	0.596	-0.376	0.209	-0.390	0.209	-0.402	0.206	-0.276
RKGFHre	0.471	0.128	0.175	0.123	0.175	0.119	0.177	0.212
RKGFHop	0.508	0.021	0.186	0.008	0.185	0.053	0.195	0.073
RKGFHru	0.508	-0.076	0.185	-0.085	0.186	-0.453	0.341	-0.065
RKGPDpr	0.208	1.340	0.240	1.356	0.241	1.342	0.240	0.958
RKGPDre	0.488	0.020	0.192	0.088	0.199	0.034	0.192	-0.031
RKGPDop	0.597	-0.408	0.183	-0.435	0.187	-0.470	0.192	-0.692
RKGPDr	0.362	0.770	0.208	0.760	0.207	0.903	0.231	0.593
RKGARpr	0.155	1.700	0.272	5.529	0.458	1.697	0.272	1.957
RKGARre	0.454	0.212	0.187	0.231	0.190	1.218	0.448	0.205
RKGARop	0.438	0.251	0.178	0.250	0.188	0.243	0.180	0.458
RKGARru	0.400	0.383	0.192	0.557	0.223	0.862	0.277	0.314
STRFHpr	0.202	1.411	0.222	1.418	0.223	1.438	0.224	1.190
STRFHre	0.215	1.285	0.214	1.291	0.215	1.423	0.221	1.477
STRFHop	0.046	3.214	0.422	3.300	0.425	3.865	0.440	4.291
STRFHru	0.262	1.078	0.225	1.087	0.226	1.670	0.276	4.477
STRPDpr	0.421	0.358	0.182	0.363	0.182	0.343	0.180	0.498
STRPDre	0.519	-0.093	0.190	-0.096	0.194	-0.106	0.190	-0.058
STRPDop	0.214	1.296	0.213	1.370	0.216	1.649	0.232	1.911
STRPDru	0.222	1.451	0.245	1.462	0.244	1.573	0.251	1.768
STRARpr	0.154	1.711	0.244	2.411	0.276	1.722	0.245	2.131
STRARre	0.341	0.678	0.192	1.109	0.228	0.681	0.193	1.411
STRARop	0.097	2.283	0.305	3.427	0.364	2.563	0.314	7.916
STRARru	0.257	1.066	0.227	2.821	0.369	3.508	0.420	1.705

Anm.: Tabellenende

B.2. Steigungsparameter

Tabelle B.2.: Steigungsparameter bzw. Faktorladungen inkl. multidim. Trennschärfe aus der 4-dimensionalen Skalierung.

Item	B_{Dim1}	B_{Dim2}	B_{Dim3}	B_{Dim4}	$MDISC_{4dims}$
BINFHpr	0.522	–	–	–	2.042
BINFHre	1.295	-1.072	–	–	2.727
BINFHop	0.404	0.005	0.344	–	0.490
BINFHru	1.145	-0.641	-1.443	2.119	0.602
BINPDpr	0.616	–	–	–	1.159
BINPDre	1.648	2.153	–	–	5.649
BINPDop	0.998	0.397	0.056	–	1.635
BINPDru	1.343	0.370	-0.692	0.659	1.199
BINARpr	0.624	–	–	–	1.486
BINARre	0.524	0.301	–	–	0.727
BINARop	0.168	0.696	1.359	–	0.442
BINARru	2.620	-0.325	2.348	5.039	0.593
DESFHpr	0.890	–	–	–	2.160
DESFHre	0.620	1.311	–	–	2.304
DESFHop	0.115	0.026	0.011	–	0.794
DESFHru	0.402	1.638	-1.484	-0.276	0.190
DESPDpr	0.454	–	–	–	1.161
DESPDre	0.605	0.780	–	–	0.752
DESPDop	0.609	0.050	0.105	–	0.674
DESPDru	1.064	-0.353	0.252	0.307	0.851
DESARpr	0.218	–	–	–	3.043
DESARre	1.350	0.431	–	–	3.157
DESARop	1.161	-0.123	0.745	–	0.655
DESARru	0.798	0.066	-0.273	0.012	0.904
DIPFHpr	1.302	–	–	–	1.780
DIPFHre	0.496	-0.472	–	–	2.532
DIPFHop	1.749	-0.034	-0.592	–	1.201
DIPFHru	0.680	0.569	0.993	-2.359	0.476
DIPPDpr	0.655	–	–	–	1.118
DIPPDre	0.333	-0.230	–	–	1.024
DIPPDop	2.144	1.448	0.720	–	1.415
DIPPDru	1.001	-0.168	0.214	-0.684	0.877
DIPARpr	1.638	–	–	–	1.863
DIPARre	3.513	-4.337	–	–	3.466
DIPARop	0.600	0.474	1.371	–	0.296
DIPARru	0.970	-0.536	0.036	0.734	1.018
RKGFHpr	0.315	–	–	–	1.920

Anm.: MDISC = multidim. Trennschärfe, Forts.: Folgeseite.

B. Aufgabenparameter

Tabelle B.2.: Steigungsparameter bzw. Faktorladungen inkl. multidim. Trennschärfe aus der 4-dimensionalen Skalierung (Fortsetzung).

Item	B_{Dim1}	B_{Dim2}	B_{Dim3}	B_{Dim4}	$MDISC_{4dims}$
RKGFHre	0.426	-0.145	–	–	0.654
RKGFHop	0.769	-0.347	-0.795	–	0.773
RKGFHru	2.243	-1.590	-3.006	-4.956	0.447
RKGPDpr	-0.078	–	–	–	3.643
RKGPDre	0.624	-0.313	–	–	0.596
RKGPDop	0.311	0.448	0.292	–	0.656
RKGPDru	1.982	0.644	2.398	0.338	1.246
RKGARpr	-0.893	–	–	–	1.101
RKGARre	0.049	2.037	–	–	1.077
RKGARop	0.418	-0.163	0.133	–	0.543
RKGARru	3.011	0.135	0.896	-1.561	1.603
STRFHpr	-0.073	–	–	–	1.655
STRFHre	-0.291	0.952	–	–	1.055
STRFHop	0.436	-0.246	2.714	–	0.782
STRFHru	0.402	-0.364	-1.379	0.805	0.237
STRPDpr	0.615	–	–	–	0.766
STRPDre	1.039	-0.245	–	–	1.330
STRPDop	-0.068	-2.719	0.849	–	0.751
STRPDru	2.066	2.380	-2.268	1.582	0.835
STRARpr	0.096	–	–	–	0.921
STRARre	0.446	-0.245	–	–	0.472
STRARop	-0.321	0.576	-0.633	–	0.287
STRARru	1.147	2.386	-2.980	1.035	0.658

Anm.: Tabellenende

C. Statistische Prüfungen der Kovariablen für die Vorstudie

Prüfung auf Varianzhomogenität und Normalverteilung

Die entsprechenden Prüfungen werden tabellarisch in den Tabellen C.1 und C.2 dargestellt und mit den Abbildungen C.1, C.2 und C.3 unterstützt.

Tabelle C.1.: Ergebnisse der Shapiro-Wilk-Tests auf Abweichung von der Normalverteilung in den Fähigkeitsschätzern der Kovariablen.

Gruppe	N	Teststatistik W		
		KFT	LGVT	Fachwissen
alle Schulen	242	0.98****	0.97****	0.91****
GYM 1	52	0.98 (n.s.)	0.89****	0.89****
GYM 2	93	0.99 (n.s.)	0.98 (n.s.)	0.84****
ISS	97	0.96***	0.97**	0.85****

Anm.: ****: $p < .001$, ***: $p < .01$, **: $p < .05$.

Tabelle C.2.: Ergebnisse der Levene-Tests auf Varianzhomogenität in den Fähigkeitsschätzern der Kovariablen zwischen den Schulen.

Variable	Teststatistik
KFT	$F(2, 242) = 2.89$ (n.s.)
LGVT	$F(2, 242) = 1.39$ (n.s.)
Fachwissen	$F(2, 242) = 1.19$ (n.s.)

Anm.: –

Gruppenvergleiche inkl. Effektstärken

Tabelle C.3 zeigt den Omnibus-Test für die schulspezifischen Unterschiedsprüfungen. Die Abbildungen C.4, C.5 und C.6 zeigen Paarvergleiche mit ****: $p < .001$, *: $p < .05$ und ns: *nicht signifikant* nach Benjamini-Hochberg-Korrektur.

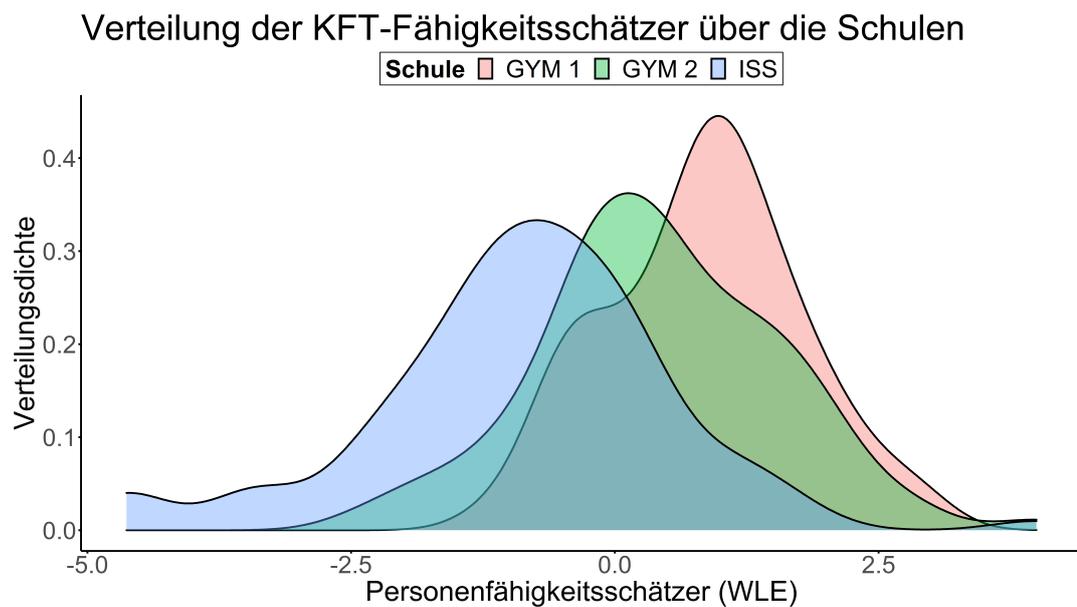


Abbildung C.1.: Grafische Verteilungsprüfung der Fähigkeitsschätzer im KFT der Vorstudie.

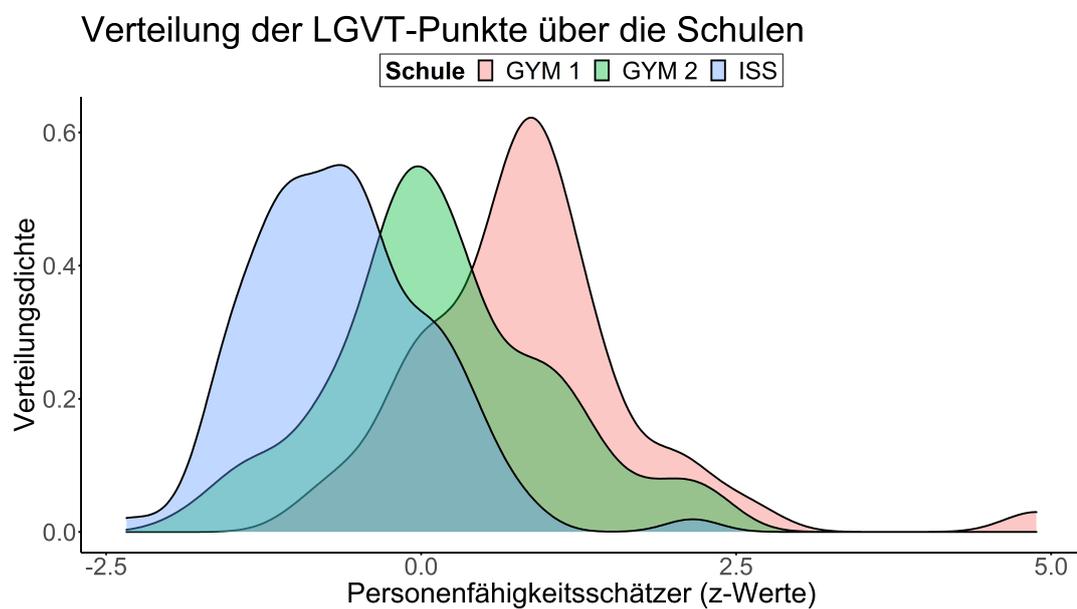


Abbildung C.2.: Grafische Verteilungsprüfung der Fähigkeitsschätzer im LGVT der Vorstudie.

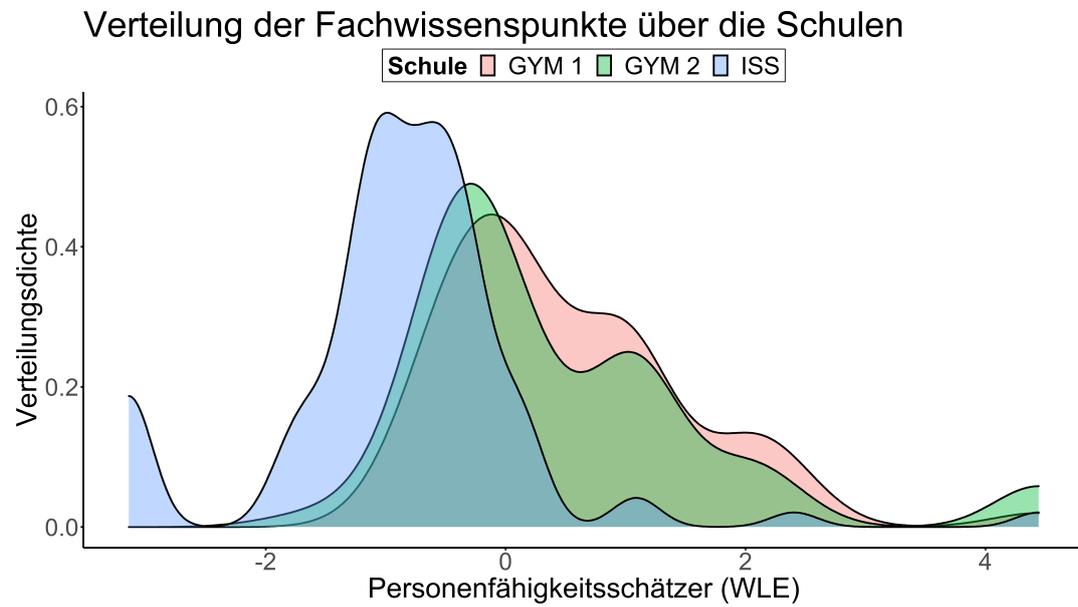


Abbildung C.3.: Grafische Verteilungsprüfung der Fähigkeitsschätzer im Fachwissen der Vorstudie.

Tabelle C.3.: Ergebnisse der Kruskal-Wallis-Tests für Schulunterschiede in der Vorstudie.

Variable	Teststatistik	Effektstärke (ϵ^2)	CI _{95%} (ϵ^2)
KFT	$H(2) = 70.90^{****}$.29	[0.20, 0.38]
LGVT	$H(2) = 87.18^{****}$.36	[0.26, 0.47]
Fachwissen	$H(2) = 89.13^{****}$.37	[0.27, 0.48]

Anm.: ****: $p < .001$, alle Tests mit $N = 242$.

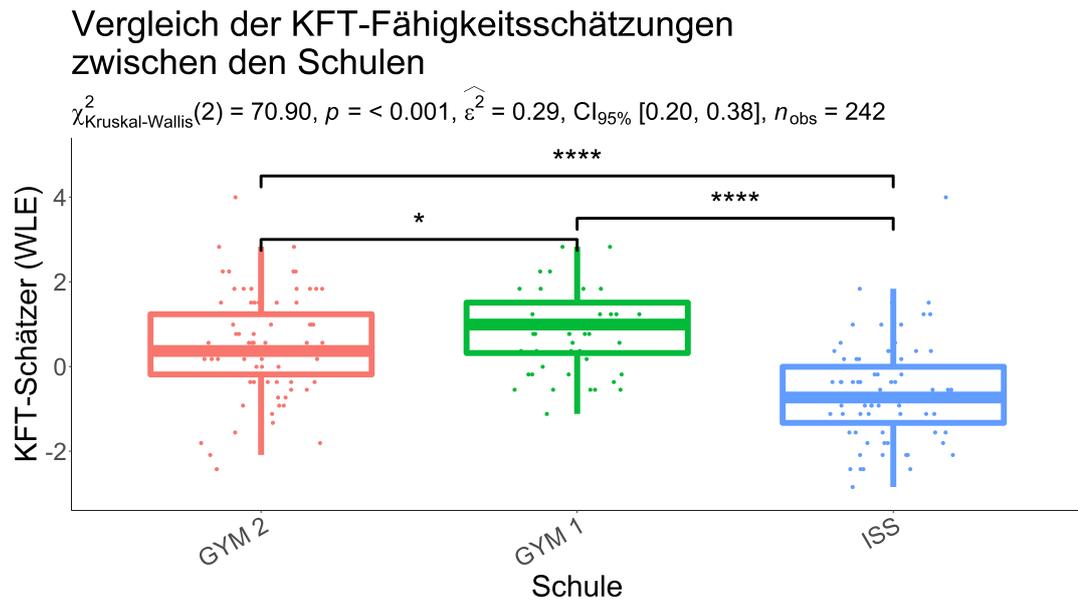


Abbildung C.4.: Schulvergleich der KFT-Fähigkeitsschätzer in der Vorstudie.

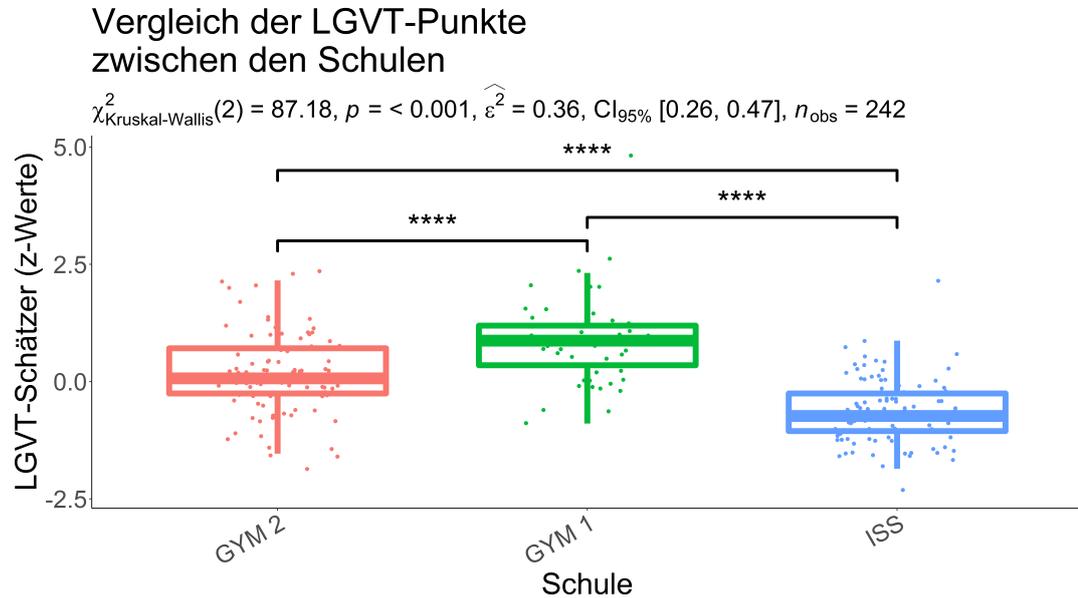


Abbildung C.5.: Schulvergleich der LGVT-Fähigkeitsschätzer in der Vorstudie.

Vergleich der FW-Fähigkeitsschätzungen zwischen den Schulen

$\chi^2_{\text{Kruskal-Wallis}}(2) = 89.13, p = < 0.001, \hat{\epsilon}^2 = 0.37, \text{CI}_{95\%} [0.27, 0.48], n_{\text{obs}} = 242$

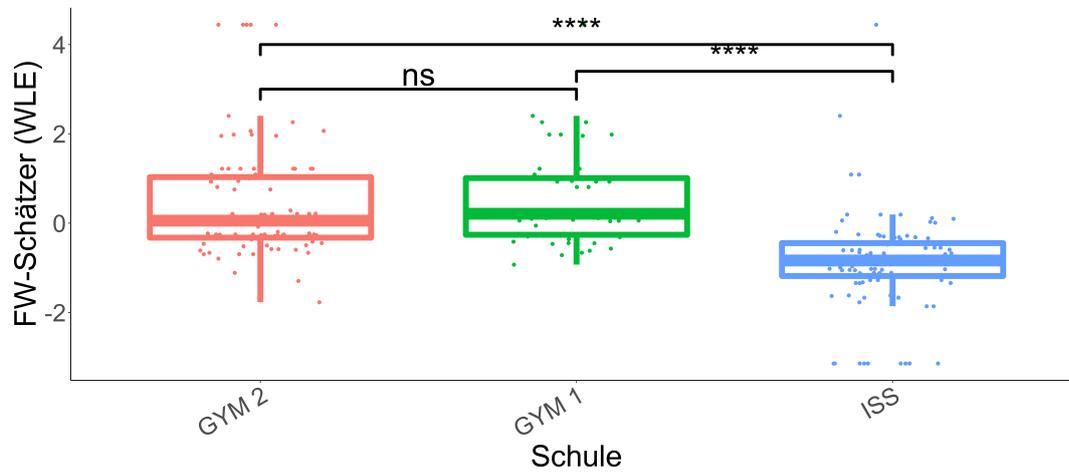


Abbildung C.6.: Schulvergleich der Fähigkeitsschätzer für das Fachwissen in der Vorstudie.

D. Zusammenhänge der Kovariablen mit den PVs der Modellaufgaben in der Vorstudie

Korrelationen der Modell-PVs mit kognitiven Fähigkeiten

Der nach Rubin (1987) gemittelte Zusammenhang zwischen der Fähigkeit zur Modellaufgabenlösung und kognitiven Fähigkeiten war signifikant (einseitig) mit $r(240) = .45, p < .001$. Abbildung D.1 zeigt die Zusammenhänge grafisch.

Korrelationen der Modell-PVs mit Fachwissen

Der nach Rubin (1987) gemittelte Zusammenhang zwischen der Fähigkeit zur Modellaufgabenlösung und Fachwissen war signifikant (einseitig) mit $r(240) = .52, p < .001$. Abbildung D.2 zeigt die Zusammenhänge grafisch.

Korrelationen der Modell-PVs mit Leseverstehen

Der nach Rubin (1987) gemittelte Zusammenhang zwischen der Fähigkeit zur Modellaufgabenlösung und Leseverstehen war signifikant (einseitig) mit $r(240) = .55, p < .001$. Abbildung D.3 zeigt die Zusammenhänge grafisch.

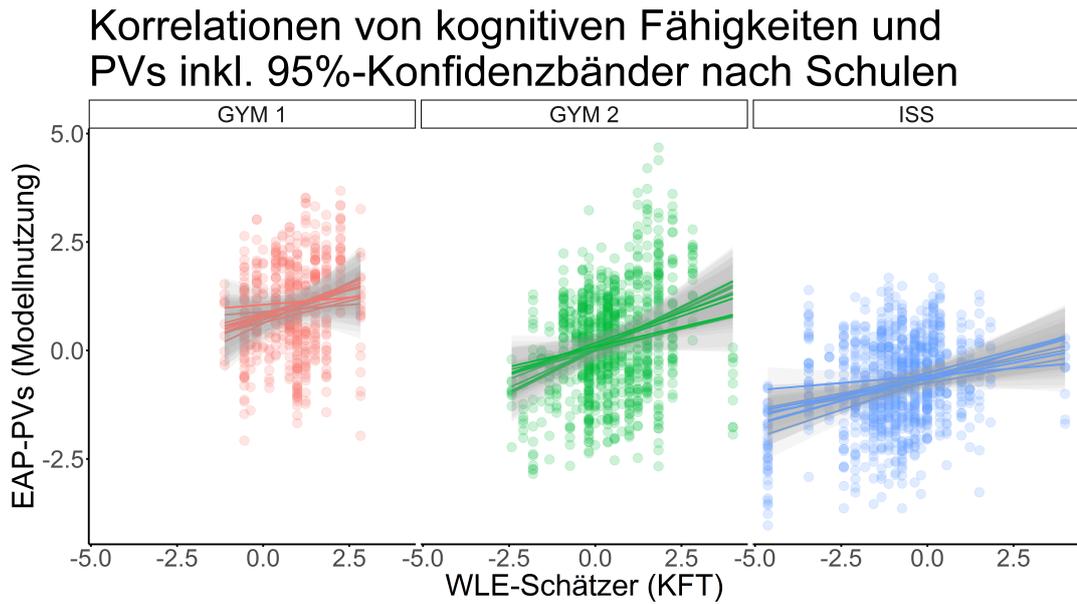


Abbildung D.1.: Regressionslinien für EAP-PVs der Modellaufgaben über den Fähigkeitsschätzern kognitiver Fähigkeiten getrennt nach Schulen.

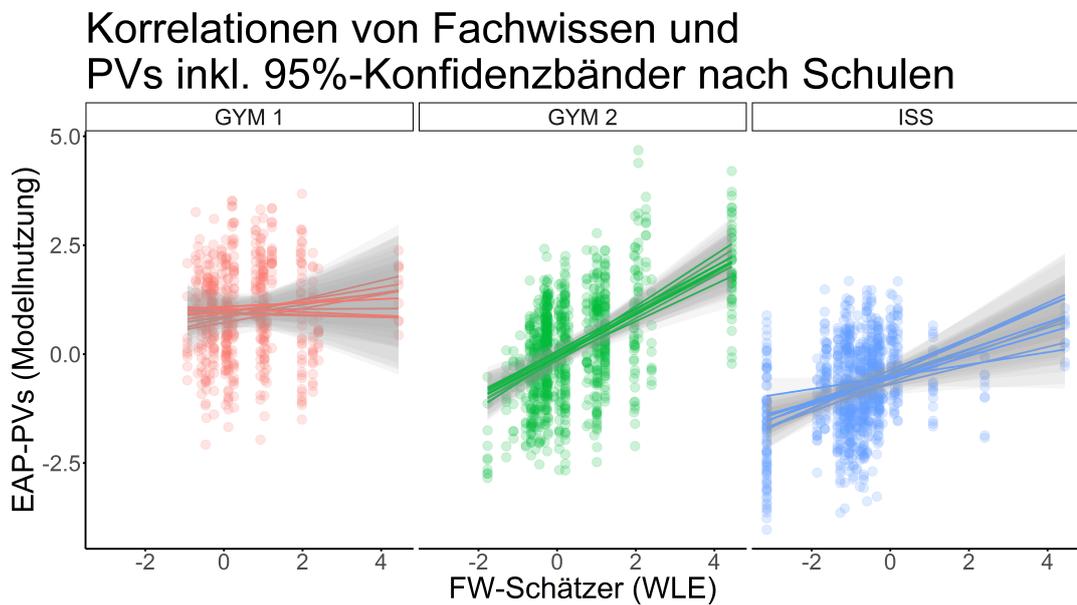


Abbildung D.2.: Regressionslinien für EAP-PVs der Modellaufgaben über den Fähigkeitsschätzern im Fachwissen getrennt nach Schulen.

Korrelationen von Leseverstehen und PVs inkl. 95%-Konfidenzbänder nach Schulen

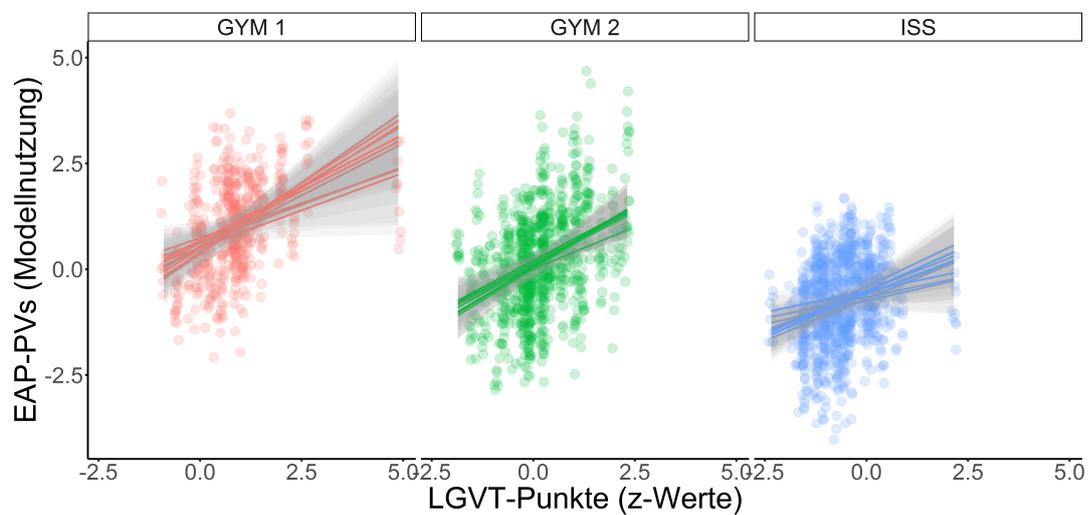


Abbildung D.3.: Regressionslinien für EAP-PVs der Modellaufgaben über den Fähigkeitsschätzern des Leseverstehens getrennt nach Schulen.

E. Schwierigkeitsvergleiche der Aufgaben in der Hauptstudie

Die Abbildungen E.1, E.2 und E.3 zeigen die Verteilungen der Schwierigkeiten unter verschiedenen Zuordnungen. Die Abbildungen E.4, E.5 und E.6 zeigen die Gruppenvergleiche der Aufgabenschwierigkeiten unter verschiedenen Zuordnungen mit **: $p < .01$ und ns oder nicht angezeigt: *nicht signifikant* nach Benjamini-Hochberg-Korrektur.

E. Schwierigkeitsvergleiche der Aufgaben in der Hauptstudie

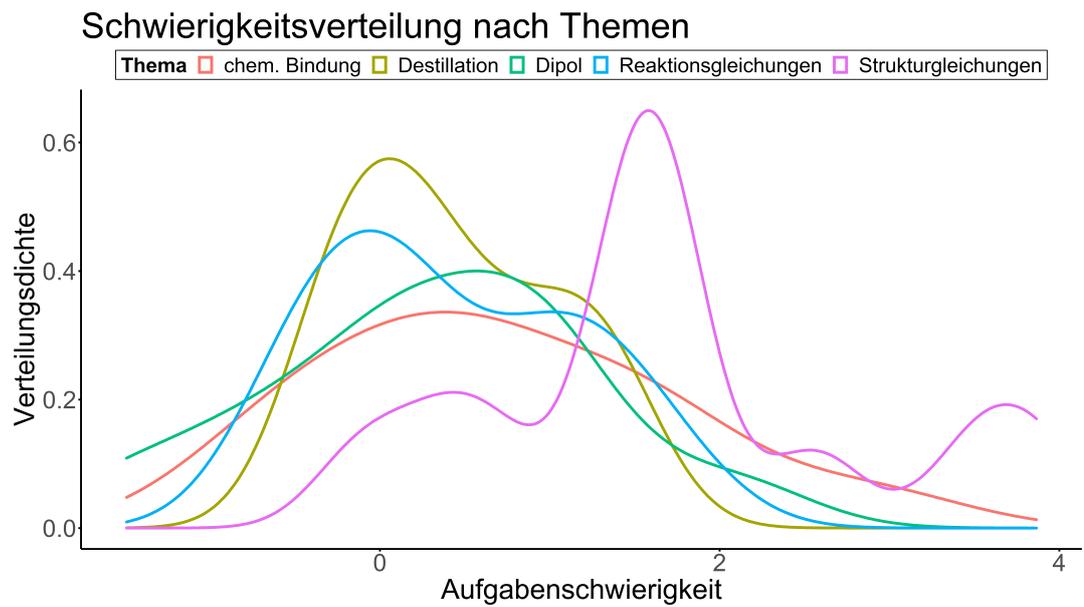


Abbildung E.1.: Schwierigkeitsverteilung aus der 4-dimensionalen Aufgabenskalierung nach Themen.

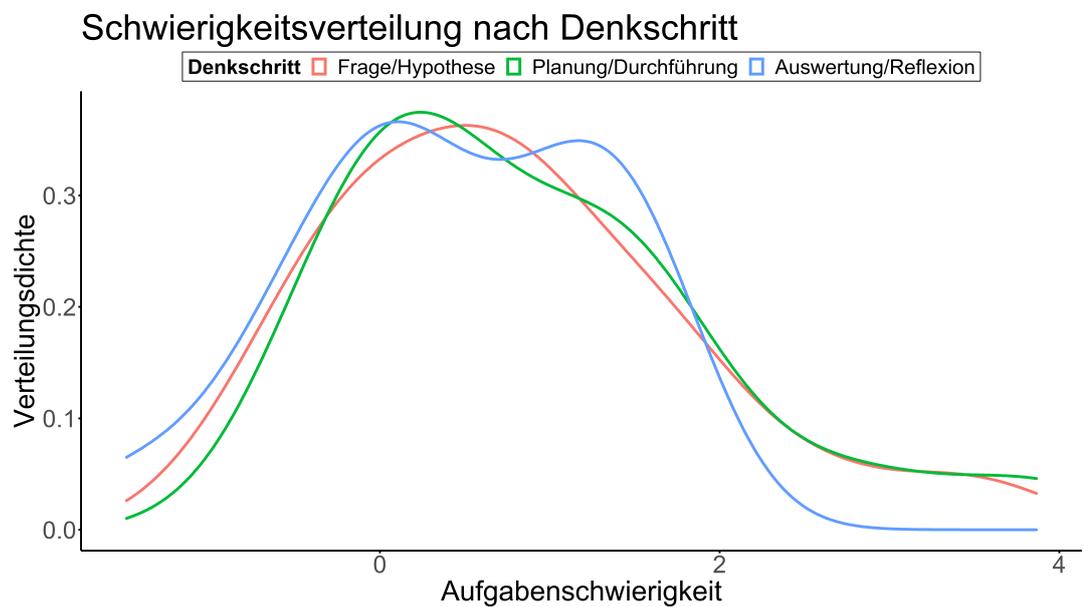


Abbildung E.2.: Schwierigkeitsverteilung aus der 4-dimensionalen Aufgabenskalierung nach Denkschritten.

Schwierigkeitsverteilung nach Modellkomponente

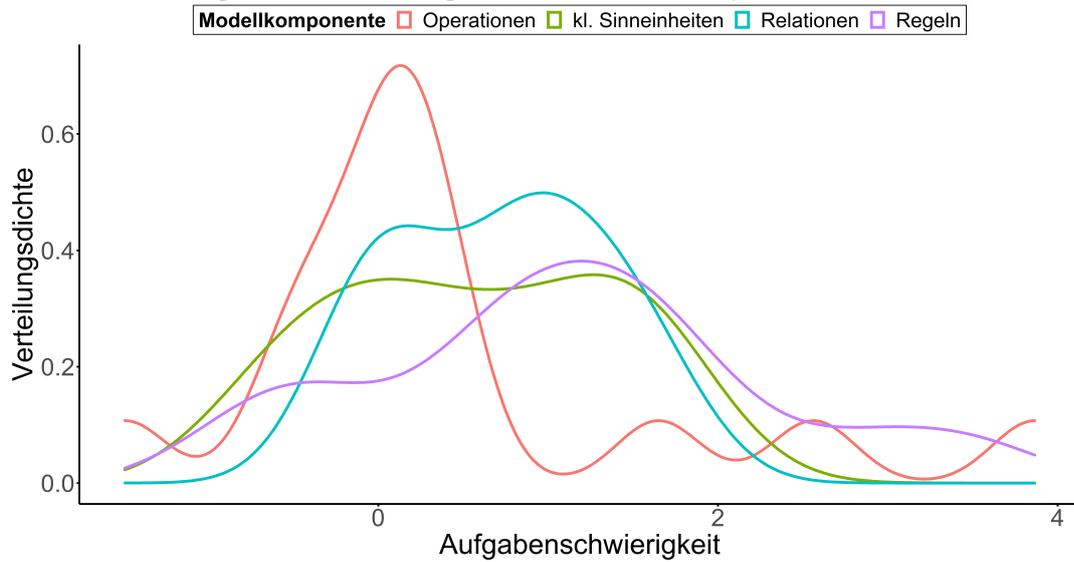


Abbildung E.3.: Schwierigkeitsverteilung aus der 4-dimensionalen Aufgabenskalierung nach Modellkomponenten.

Vergleich der Aufgabenschwierigkeiten zwischen den Themen

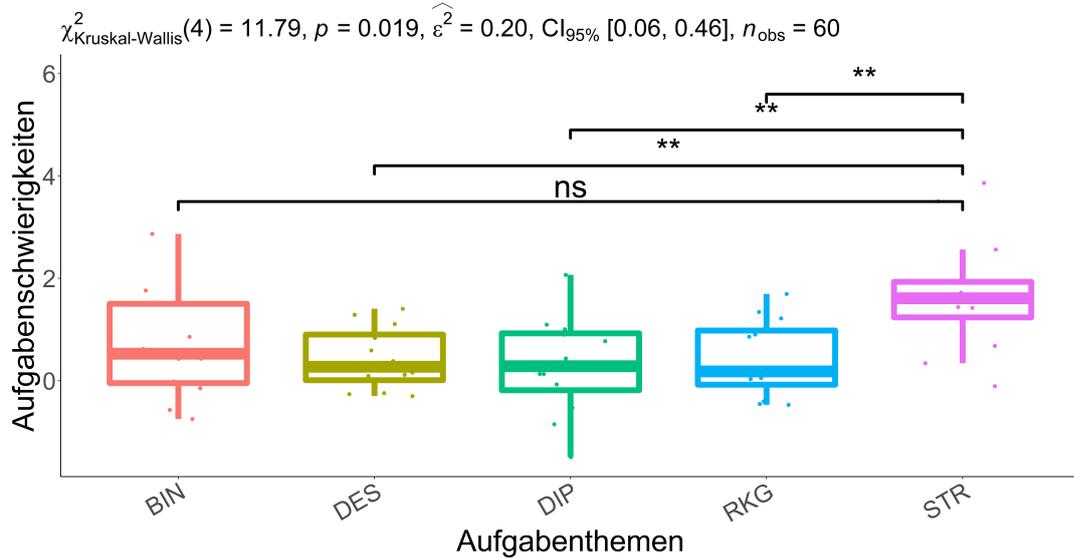


Abbildung E.4.: Schwierigkeitsvergleiche aus der 4-dimensionalen Aufgabenskalierung nach Themen.

E. Schwierigkeitsvergleiche der Aufgaben in der Hauptstudie

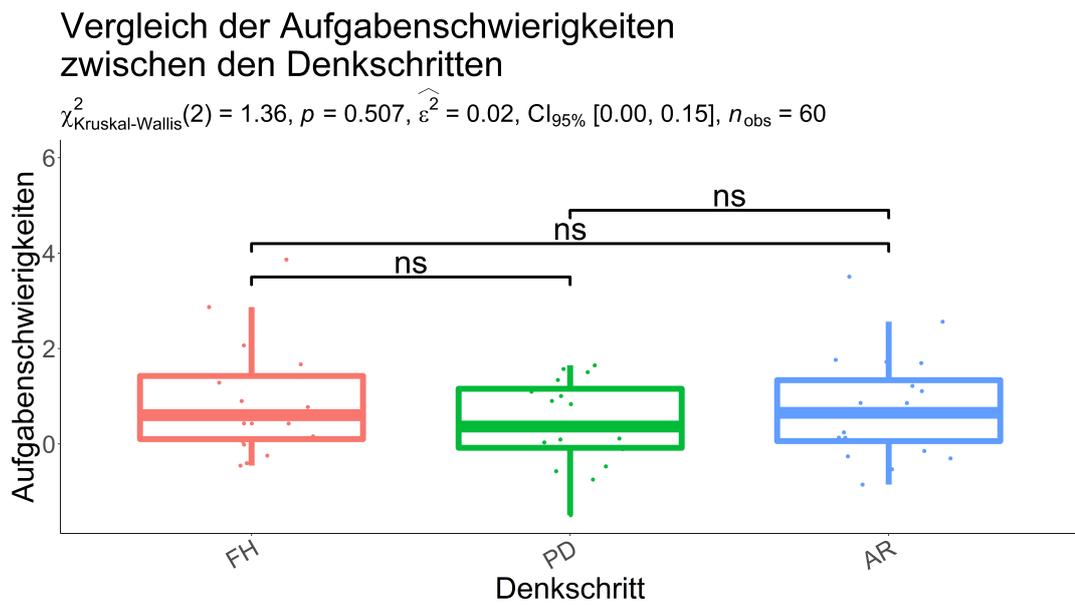


Abbildung E.5.: Schwierigkeitsvergleiche aus der 4-dimensionalen Aufgabenskalierung nach Denkschritten.

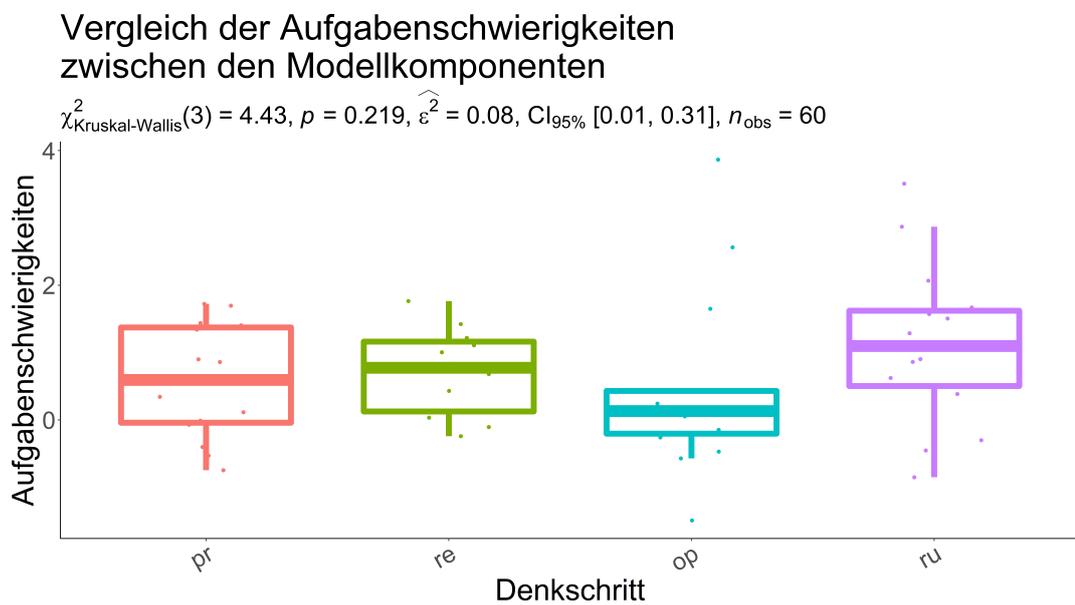


Abbildung E.6.: Schwierigkeitsvergleiche aus der 4-dimensionalen Aufgabenskalierung nach Modellkomponenten.

F. Konstruktionsmanual

Modellzelle: FH_pr	Fragestellungen und Hypothese zu Merkmalen und Zusammenhängen mit Hilfe der kl. Sinneinheiten von Modellen entwickeln.
Modellbestandteil: kl. Sinneinheiten	Denkschritt: Fragestellung und Hypothese
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell
Frage	Welche Vermutung können sie mit Hilfe des Modells aufstellen?
Antwortmöglichkeiten	<p>Eine Hypothese die widerspruchsfrei auf Basis des Itemstamms und der kl. Sinneinheiten des Modells abgeleitet werden kann.</p> <p>Drei Hypothesen, die nicht widerspruchsfrei auf Basis des Itemstamms und der kl. Sinneinheiten des Modells abgeleitet werden können.</p>
gemessene Fähigkeit	Zuordnung einer Hypothese zu gegebenen Informationen auf Basis der kl. Sinneinheiten eines chemiespezifischen Modells.

Modellzelle: FH_re	Fragestellungen und Hypothese zu Merkmalen und Zusammenhängen mit Hilfe der Relationen von Modellen entwickeln.
Modellbestandteil: Relationen	Denkschritt: Fragestellung und Hypothese
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell
Frage	Welche Vermutung können sie mit Hilfe des Modells aufstellen?
Antwortmöglichkeiten	<p>Eine Hypothese die widerspruchsfrei auf Basis des Itemstamms und der Relationen des Modells abgeleitet werden kann.</p> <p>Drei Hypothesen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Relationen des Modells abgeleitet werden können.</p>
gemessene Fähigkeit	Zuordnung einer Hypothese zu gegebenen Informationen auf Basis der Relationen eines chemiespezifischen Modells.

Modellzelle: FH_op	Fragestellungen und Hypothese zu Merkmalen und Zusammenhängen mit Hilfe der Operationen von Modellen entwickeln.
Modellbestandteil: Operationen	Denkschritt: Fragestellung und Hypothese
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell
Frage	Welche Vermutung können sie mit Hilfe des Modells aufstellen?
Antwortmöglichkeiten	Eine Hypothese die widerspruchsfrei auf Basis des Itemstamms und der Operationen des Modells abgeleitet werden kann. Drei Hypothesen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Operationen des Modells abgeleitet werden können.
gemessene Fähigkeit	Zuordnung einer Hypothese zu gegebenen Informationen auf Basis der Operationen eines chemiespezifischen Modells.

Modellzelle: FH_ru	Fragestellungen und Hypothese zu Merkmalen und Zusammenhängen mit Hilfe der Regeln von Modellen entwickeln.
Modellbestandteil: Regeln	Denkschritt: Fragestellung und Hypothese
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell
Frage	Welche Vermutung können sie mit Hilfe des Modells aufstellen?
Antwortmöglichkeiten	Eine Hypothese die widerspruchsfrei auf Basis des Itemstamms und der Regeln des Modells abgeleitet werden kann. Drei Hypothesen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Regeln des Modells abgeleitet werden können.
gemessene Fähigkeit	Zuordnung einer Hypothese zu gegebenen Informationen auf Basis der Regeln eines chemiespezifischen Modells.

Modellzelle: PD_el	Eine hypothesenbezogene Untersuchung mit den kl. Sinneinheiten am Modell planen.
Modellbestandteil: kl. Sinneinheiten	Denkschritt: Planung und Durchführung
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell 4. Eine Vermutung
Frage	Welche Untersuchung können sie mit Hilfe des Modells planen?
Antwortmöglichkeiten	<p>Eine Untersuchung die widerspruchsfrei auf Basis des Itemstamms und der kl. Sinneinheiten des Modells abgeleitet werden kann.</p> <p>Drei Untersuchungen, die nicht widerspruchsfrei auf Basis des Itemstamms und der kl. Sinneinheiten des Modells abgeleitet werden können.</p>
gemessene Fähigkeit	Zuordnung einer Untersuchung zu gegebenen Informationen auf Basis der kl. Sinneinheiten eines chemiespezifischen Modells.

Modellzelle: PD_re	Eine hypothesenbezogene Untersuchung mit den Relationen am Modell planen.
Modellbestandteil: Relationen	Denkschritt: Planung und Durchführung
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell 4. Eine Vermutung
Frage	Welche Untersuchung können sie mit Hilfe des Modells planen?
Antwortmöglichkeiten	<p>Eine Untersuchung die widerspruchsfrei auf Basis des Itemstamms und der Relationen des Modells abgeleitet werden kann.</p> <p>Drei Untersuchungen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Relationen des Modells abgeleitet werden können.</p>
gemessene Fähigkeit	Zuordnung einer Untersuchung zu gegebenen Informationen auf Basis der Relationen eines chemiespezifischen Modells.

Modellzelle: PD_op	Eine hypothesenbezogene Untersuchung mit den Operationen am Modell planen.
Modellbestandteil: Operationen	Denkschritt: Planung und Durchführung
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell 4. Eine Vermutung
Frage	Welche Untersuchung können sie mit Hilfe des Modells planen?
Antwortmöglichkeiten	Eine Untersuchung die widerspruchsfrei auf Basis des Itemstamms und der Operationen des Modells abgeleitet werden kann. Drei Untersuchungen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Operationen des Modells abgeleitet werden können.
gemessene Fähigkeit	Zuordnung einer Untersuchung zu gegebenen Informationen auf Basis der Operationen eines chemiespezifischen Modells.

Modellzelle: PD_ru	Eine hypothesenbezogene Untersuchung mit den Regeln am Modell planen.
Modellbestandteil: Regeln	Denkschritt: Planung und Durchführung
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell 4. Eine Vermutung
Frage	Welche Untersuchung können sie mit Hilfe des Modells planen?
Antwortmöglichkeiten	Eine Untersuchung die widerspruchsfrei auf Basis des Itemstamms und der Regeln des Modells abgeleitet werden kann. Drei Untersuchungen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Regeln des Modells abgeleitet werden können.
gemessene Fähigkeit	Zuordnung einer Untersuchung zu gegebenen Informationen auf Basis der Regeln eines chemiespezifischen Modells.

Modellzelle: AR_el	Gewonnene Daten mit Hilfe der kl. Sinneinheiten eines Modells hypothesenbezogen interpretieren.
Modellbestandteil: kl. Sinneinheiten	Denkschritt: Auswertung und Reflexion
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell 4. Eine Vermutung 5. Am Modell gewonnene Daten
Frage	Welche Schlussfolgerungen können sie mit Hilfe des Modells ziehen?
Antwortmöglichkeiten	<p>Eine Schlussfolgerung die widerspruchsfrei auf Basis des Itemstamms und der kl. Sinneinheiten des Modells abgeleitet werden kann.</p> <p>Drei Schlussfolgerungen, die nicht widerspruchsfrei auf Basis des Itemstamms und der kl. Sinneinheiten des Modells abgeleitet werden können.</p>
gemessene Fähigkeit	Zuordnung einer Schlussfolgerung zu gegebenen Daten auf Basis der kl. Sinneinheiten eines chemiespezifischen Modells.

Modellzelle: AR_re	Gewonnene Daten mit Hilfe der Relationen eines Modells hypothesenbezogen interpretieren.
Modellbestandteil: Relationen	Denkschritt: Auswertung und Reflexion
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell 4. Eine Vermutung 5. Am Modell gewonnene Daten
Frage	Welche Schlussfolgerungen können sie mit Hilfe des Modells ziehen?
Antwortmöglichkeiten	<p>Eine Schlussfolgerung die widerspruchsfrei auf Basis des Itemstamms und der Relationen des Modells abgeleitet werden kann.</p> <p>Drei Schlussfolgerungen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Relationen des Modells abgeleitet werden können.</p>
gemessene Fähigkeit	Zuordnung einer Schlussfolgerung zu gegebenen Daten auf Basis der Relationen eines chemiespezifischen Modells.

Modellzelle: AR_op	Gewonnene Daten mit Hilfe der Operationen eines Modells hypothesenbezogen interpretieren.
Modellbestandteil: Operationen	Denkschritt: Auswertung und Reflexion
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell 4. Eine Vermutung 5. Am Modell gewonnene Daten
Frage	Welche Schlussfolgerungen können sie mit Hilfe des Modells ziehen?
Antwortmöglichkeiten	<p>Eine Schlussfolgerung die widerspruchsfrei auf Basis des Itemstamms und der Operationen des Modells abgeleitet werden kann.</p> <p>Drei Schlussfolgerungen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Operationen des Modells abgeleitet werden können.</p>
gemessene Fähigkeit	Zuordnung einer Schlussfolgerung zu gegebenen Daten auf Basis der Operationen eines chemiespezifischen Modells.

Modellzelle: AR_ru	Gewonnene Daten mit Hilfe der Regeln eines Modells hypothesenbezogen interpretieren.
Modellbestandteil: Regeln	Denkschritt: Auswertung und Reflexion
Itemstruktur	Iteminhalt
Stamm	<ol style="list-style-type: none"> 1. Fachwissen 2. Fachwissen für plausible Distraktoren 3. Ein Modell 4. Eine Vermutung 5. Am Modell gewonnene Daten
Frage	Welche Schlussfolgerungen können sie mit Hilfe des Modells ziehen?
Antwortmöglichkeiten	<p>Eine Schlussfolgerung die widerspruchsfrei auf Basis des Itemstamms und der Regeln des Modells abgeleitet werden kann.</p> <p>Drei Schlussfolgerungen, die nicht widerspruchsfrei auf Basis des Itemstamms und der Regeln des Modells abgeleitet werden können.</p>
gemessene Fähigkeit	Zuordnung einer Schlussfolgerung zu gegebenen Daten auf Basis der Regeln eines chemiespezifischen Modells.

G. Testheftdesign

	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8	Frage 9	Frage 10	Frage 11	Frage 12	Frage 13	Frage 14	Frage 15	
R1	Bin_FH_ru	Bin_FH_pr	Rkg_FH_op	Dip_AR_op	Dip_AR_pr	Str_PD_re	Des_PD_re	Str_PD_op	Rkg_FH_re	Rkg_FH_re	Des_PD_ru	Dip_AR_ru	Bin_FH_op	Bin_FH_re	Str_PD_pr	Des_PD_pr
R2	Str_PD_re	Des_PD_re	Str_PD_op	Rkg_FH_op	Dip_AR_op	Dip_AR_pr	Rkg_FH_ru	Bin_FH_op	Bin_FH_pr	Rkg_FH_re	Str_PD_op	Des_PD_re	Str_PD_op	Rkg_FH_re	Dip_AR_op	Dip_AR_pr
R3	Dip_AR_ru	Bin_FH_op	Bin_FH_re	Str_PD_pr	Des_PD_pr	Rkg_FH_ru	Bin_FH_pr	Rkg_FH_re	Dip_AR_op	Dip_AR_pr	Str_PD_re	Des_PD_re	Str_PD_op	Rkg_FH_re	Rkg_FH_re	Des_PD_ru
R4	Rkg_PD_op	Str_AR_op	Des_AR_re	Rkg_PD_re	Str_AR_pr	Des_AR_ru	Bin_PD_re	Dip_FH_ru	Des_AR_pr	Dip_FH_pr	Rkg_PD_op	Dip_FH_op	Bin_PD_op	Bin_PD_pr	Str_AR_re	Str_AR_pr
R5	Des_AR_ru	Bin_PD_re	Dip_FH_ru	Des_AR_pr	Dip_FH_pr	Rkg_PD_op	Str_AR_op	Des_AR_re	Rkg_PD_re	Str_AR_pr	Des_AR_ru	Bin_PD_re	Dip_FH_ru	Des_AR_pr	Dip_FH_pr	Dip_FH_pr
R6	Rkg_PD_ru	Dip_FH_op	Bin_PD_op	Bin_PD_pr	Str_AR_re	Rkg_PD_op	Str_AR_op	Des_AR_re	Rkg_PD_re	Str_AR_pr	Des_AR_ru	Bin_PD_re	Dip_FH_ru	Des_AR_pr	Dip_FH_pr	Dip_FH_pr
R7	Des_FH_re	Rkg_AR_re	Dip_PD_ru	Des_FH_ru	Bin_AR_pr	Dip_PD_op	Dip_PD_pr	Str_FH_pr	Rkg_AR_ru	Bin_AR_re	Str_FH_re	Bin_AR_op	Des_FH_pr	Rkg_AR_op	Str_FH_op	Str_FH_op
R8	Dip_PD_op	Dip_PD_pr	Str_FH_pr	Rkg_AR_ru	Bin_AR_re	Str_FH_re	Bin_AR_op	Des_FH_pr	Rkg_AR_op	Str_FH_op	Des_FH_re	Rkg_AR_re	Dip_PD_ru	Des_FH_ru	Bin_AR_pr	Bin_AR_pr
R9	Str_FH_re	Bin_AR_op	Des_FH_pr	Rkg_AR_op	Str_FH_op	Des_FH_re	Rkg_AR_re	Dip_PD_ru	Des_FH_ru	Bin_AR_pr	Dip_PD_op	Dip_PD_pr	Str_FH_pr	Rkg_AR_ru	Bin_AR_re	Bin_AR_re
DW1	Bin_FH_pr	Bin_FH_re	Bin_FH_ru	Rkg_FH_pr	Rkg_FH_re	Rkg_FH_op	Str_PD_pr	Str_PD_re	Str_PD_ru	Des_PD_pr	Des_PD_re	Des_PD_ru	Dip_AR_re	Dip_AR_op	Dip_AR_ru	Dip_AR_ru
DW2	Dip_AR_re	Dip_AR_op	Dip_AR_ru	Des_PD_pr	Des_PD_re	Des_PD_ru	Str_PD_pr	Str_PD_re	Str_PD_ru	Bin_FH_pr	Bin_FH_re	Bin_FH_ru	Rkg_FH_pr	Rkg_FH_re	Rkg_FH_ru	Rkg_FH_op
DW3	Bin_PD_pr	Bin_PD_re	Bin_PD_ru	Rkg_PD_pr	Rkg_PD_re	Rkg_PD_op	Des_AR_pr	Des_AR_re	Des_AR_ru	Str_AR_pr	Str_AR_re	Str_AR_ru	Dip_FH_re	Dip_FH_op	Dip_FH_ru	Dip_FH_ru
DW4	Dip_FH_re	Dip_FH_op	Dip_FH_ru	Str_AR_pr	Str_AR_re	Str_AR_ru	Des_AR_pr	Des_AR_re	Des_AR_ru	Rkg_PD_pr	Rkg_PD_re	Rkg_PD_ru	Bin_PD_pr	Bin_PD_re	Bin_PD_ru	Bin_PD_ru
DW5	Bin_AR_pr	Bin_AR_re	Bin_AR_ru	Des_FH_pr	Des_FH_re	Des_FH_op	Dip_PD_re	Dip_PD_ru	Dip_PD_ru	Rkg_AR_pr	Rkg_AR_re	Rkg_AR_ru	Str_FH_re	Str_FH_op	Str_FH_ru	Str_FH_ru
DW6	Str_FH_pr	Str_FH_re	Str_FH_ru	Rkg_AR_pr	Rkg_AR_re	Rkg_AR_op	Dip_PD_re	Dip_PD_ru	Dip_PD_ru	Des_FH_pr	Des_FH_re	Des_FH_ru	Bin_AR_pr	Bin_AR_re	Bin_AR_ru	Bin_AR_ru
MK1	Bin_FH_pr	Dip_PD_pr	Rkg_FH_pr	Str_AR_pr	Des_AR_re	Dip_PD_re	Rkg_FH_re	Bin_FH_op	Des_AR_op	Dip_PD_op	Str_AR_op	Bin_FH_ru	Des_AR_ru	Rkg_FH_ru	Str_AR_ru	Str_AR_ru
MK2	Bin_FH_ru	Des_AR_ru	Rkg_FH_ru	Str_AR_ru	Bin_FH_op	Des_AR_op	Dip_PD_op	Str_AR_op	Des_AR_re	Dip_PD_re	Rkg_FH_re	Bin_FH_pr	Dip_PD_pr	Rkg_FH_pr	Str_AR_pr	Str_AR_pr
MK3	Des_FH_re	Dip_AR_re	Rkg_PD_re	Bin_PD_op	Des_FH_op	Dip_AR_op	Str_FH_op	Bin_PD_op	Des_FH_ru	Rkg_PD_ru	Str_FH_ru	Bin_PD_pr	Dip_AR_pr	Rkg_PD_pr	Str_FH_pr	Str_FH_pr
MK4	Bin_PD_pr	Dip_AR_pr	Rkg_PD_pr	Str_FH_pr	Bin_PD_ru	Des_FH_ru	Rkg_PD_ru	Str_FH_ru	Bin_PD_op	Des_FH_op	Dip_AR_op	Str_FH_op	Des_FH_re	Dip_AR_re	Rkg_PD_re	Rkg_PD_re
MK5	Bin_AR_op	Des_PD_op	Dip_FH_op	Str_PD_op	Bin_AR_ru	Des_PD_ru	Rkg_AR_ru	Str_PD_ru	Bin_AR_pr	Rkg_AR_pr	Str_PD_pr	Bin_AR_op	Des_PD_op	Dip_FH_op	Dip_FH_op	Dip_FH_op
MK6	Des_PD_re	Dip_FH_re	Rkg_AR_re	Bin_AR_pr	Dip_FH_pr	Rkg_AR_ru	Str_PD_pr	Bin_AR_ru	Des_PD_ru	Rkg_AR_ru	Str_PD_ru	Bin_AR_op	Des_PD_op	Dip_FH_op	Str_PD_op	Str_PD_op
MK7	Bin_FH_ru	Dip_PD_ru	Rkg_FH_ru	Str_AR_ru	Des_AR_pr	Dip_PD_pr	Rkg_FH_pr	Bin_FH_re	Des_AR_re	Dip_PD_re	Str_AR_re	Bin_FH_op	Des_AR_op	Rkg_FH_op	Str_AR_op	Str_AR_op
MK8	Bin_FH_op	Des_AR_op	Rkg_FH_op	Str_AR_op	Bin_FH_re	Des_AR_re	Dip_PD_re	Str_AR_re	Des_AR_pr	Dip_PD_pr	Rkg_FH_pr	Bin_FH_ru	Dip_PD_ru	Rkg_FH_ru	Str_AR_ru	Str_AR_ru
MK9	Des_FH_pr	Dip_AR_pr	Rkg_PD_pr	Bin_PD_op	Bin_PD_pr	Str_FH_re	Bin_PD_re	Des_FH_re	Des_FH_ru	Dip_AR_re	Str_FH_re	Bin_PD_op	Des_FH_op	Rkg_PD_op	Str_FH_op	Str_FH_op
MK10	Bin_PD_op	Des_FH_op	Rkg_PD_op	Str_FH_op	Bin_PD_re	Des_FH_re	Dip_AR_re	Str_FH_re	Bin_PD_ru	Dip_AR_ru	Rkg_PD_ru	Str_FH_ru	Des_FH_pr	Dip_AR_pr	Rkg_PD_pr	Rkg_PD_pr
MK11	Bin_AR_ru	Dip_FH_ru	Rkg_AR_ru	Str_PD_ru	Des_PD_pr	Dip_FH_pr	Rkg_AR_pr	Bin_AR_op	Des_PD_op	Rkg_AR_op	Str_PD_op	Bin_AR_re	Des_PD_re	Dip_FH_re	Str_PD_re	Str_PD_re
MK12	Bin_AR_re	Des_PD_re	Dip_FH_re	Str_PD_re	Bin_AR_op	Des_PD_op	Rkg_AR_op	Str_PD_op	Des_PD_pr	Dip_FH_pr	Rkg_AR_pr	Bin_AR_ru	Dip_FH_ru	Rkg_AR_ru	Str_PD_ru	Str_PD_ru

H. Beispielaufgaben

BN_PD_pfr

Hintergrundinformationen

In dieser Aufgabe geht es um Modelle von Kristallen. Man kann davon ausgehen, dass sich die einzelnen Teilchen eines Kristalls wie kleine Kugeln verhalten.

Kochsalz ist eine aus dem Alltag bekannte Verbindung. Sie bildet Kristalle aus Ionenbindungen, die ein regelmäßiges Gitter aufweisen. Abbildung 1 zeigt die Struktur eines solchen Kristallgitters im Querschnitt. Die einzelnen Teilchen des Gitters sind positiv geladene Kationen und negativ geladene Anionen. Diese Gitterteilchen haben unterschiedliche Radien und ziehen sich durch die entgegengesetzten Ladungen an.

Kupfer ist ein Metall. Metalle bilden Kristalle aus Metallbindungen, die ein regelmäßiges Gitter aufweisen. Abbildung 2 zeigt die Struktur eines solchen Kristallgitters im Querschnitt. Die einzelnen Teilchen des Gitters sind Metallatome. Diese Gitterteilchen haben die gleichen Radien und werden von frei beweglichen Elektronen zusammengehalten.

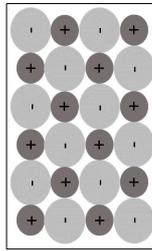


Abbildung 1: Ein Subkristallgitter im Querschnitt. Die Kationen sind durch ein ... gekennzeichnet. Die Anionen sind durch ein ... gekennzeichnet.

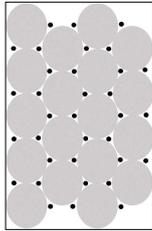


Abbildung 2: Ein Metallgitter im Querschnitt. Die Elektronen sind durch die schwarzen Punkte gekennzeichnet.

BN_PD_pfr

Aufgabe

Die Materialwissenschaftler Claudia und Patrick wurden damit beauftragt, die Eigenschaften von Salz- und Metallkristallen zu untersuchen, wenn eine Kraft auf sie einwirkt. Die beiden meinen, dass es dafür wichtig ist, auf bestimmte **Einzelteile der Modelle** zu achten. Die Einzelteile sind in diesem Fall die **kugelförmigen Gitterbestandteile**.

Sie vermuten außerdem: Die Gitterbestandteile verschieben sich gegenseitig, wenn eine Kraft auf den Kristall einwirkt.

Welches Vorgehen können sie mit Hilfe der Gittermodelle und ihrer Vermutung für den Versuch planen? **Kreuzen an!**

Sie können ein Vorgehen planen bei dem sie vergleichen was passiert, wenn sie versuchen mit einer Zange ...

- ... zwei verschiedene Salzkristalle zu verformen.
- ... einen Salzkristall und eine Metallmünze zu verformen.
- ... zwei Münzen aus verschiedenen Metallen zu verformen.
- ... einen Kupferstab und eine Kupfermünze zu verformen.

H. Beispielaufgaben

DIP_AR_op

Aufgabe

Marco und Frank sind Schüler eines Leistungskurses Chemie. Sie müssen für ihre mündliche Abiturprüfung die Eigenschaften von Wasser kennen. Sie wissen, dass das Thema Löslichkeit dabei eine wichtige Rolle spielen wird. Die beiden meinen, dass es dafür wichtig ist auf bestimmte **Veränderungen zwischen den Einzelteilen des Modells** zu achten. Die Veränderung ist in diesem Fall die **Verschiebung von Elektronen**. Sie haben vermutet: Je stärker die Verschiebung von Elektronen innerhalb eines Moleküls ist, desto besser ist das Molekül wasserlöslich. Die beiden haben versucht 3 verschiedene Stoffe (A, B, C) in Wasser zu lösen. Dabei wurden folgende Ergebnisse notiert:

Stoff	Moleküle tragen Partialladung	Moleküle sind gewinkelt	Moleküle besitzen ähnliche Anzahl Sauerstoffatome wie Wasser	Moleküle tragen viele freie Elektronenpaare	Stoff löst sich in Wasser
A	Ja	Nein	Nein	Nein	Ja
B	Nein	Ja	Ja	Ja	Nein
C	Ja	Ja	Ja	Ja	Ja

In Bezug auf ihre Vermutung und den Versuch: Welche Schlussfolgerung können sie mit Hilfe des Modells ziehen? **Kreuze an!**

Ihre Vermutung stimmt, weil die Stoffe ...

- ... B und C freie Elektronenpaare tragen.
- ... A und B sich in mehr als zwei Eigenschaften unterscheiden.
- ... A, B und C sich nur bei den Partialladungen unterscheiden.
- ... A und C in Wasser löslich sind.

DIP_AR_op

Hintergrundinformationen

Das gewinkelte Wassermolekül trägt sogenannte Partialladungen. Aus diesem Grund wird es auch als Dipolmolekül bezeichnet. Die Partialladungen bilden sich, weil die negativ geladenen Bindungselektronen stärker vom Sauerstoffatom angezogen werden. So erhöht sich dort die Dichte an negativer Ladung. Das Maß für die Anziehung von Elektronen in einer Verbindung nennt man Elektronegativität. Außerdem befinden sich zwei freie Elektronenpaare am Sauerstoffatom. Man kann annehmen, dass Wassermoleküle sich wie kleine Magneten verhalten (Abbildung 1), weil sie sich über die gebildeten Partialladungen gegenseitig anziehen. Aus diesem Grund hat Wasser auch einen relativ hohen Siedepunkt.

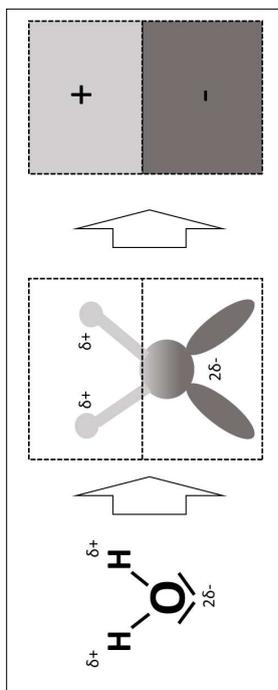


Abbildung 1: Die Partialladungen (δ^+ und δ^-) im Wassermolekül sind sehr deutlich voneinander getrennt. Aus diesem Grund kann das ganze Molekül als kleiner Stabmagnet betrachtet werden.

Hintergrundinformationen

In einer chemischen Reaktion reagieren unzählige Teilchen (Atome oder Moleküle) miteinander. Reaktionsgleichungen können als Modelle für chemische Reaktionen verstanden werden. Jedem chemischen Element werden dazu Symbole aus dem Periodensystem zugeordnet. Die Symbole der Ausgangsstoffe (Edukte) stehen links vom Reaktionspfeil und die Symbole der Reaktionsprodukte stehen rechts vom Reaktionspfeil. Mit Zahlen werden die Mengenverhältnisse der Elemente in der Reaktion dargestellt. Auf diese Weise können komplizierte Vorgänge vereinfacht beschrieben werden. Auch die Vorhersage von physikalischen Größen wie Masse oder Volumen der Produkte ist damit möglich. Folgend steht eine allgemeine Reaktionsgleichung. Die Symbole werden in Tabelle 1 erklärt.



Tabelle 1: Bedeutung der einzelnen Symbole in einer Reaktionsgleichung.

Symbol	Bedeutung
Elementsymbole „A“, „B“, „C“	Die Symbole stehen jeweils für ein Element aus dem Periodensystem.
Verknüpfungssymbol „+“	Das „Plus“ steht für „reagiert mit“ und trennt jeweils die Symbole auf der Produktseite (rechts vom „Reaktionspfeil“) und die Symbole auf der Eduktseite (links vom „Reaktionspfeil“).
Reaktionspfeil „→“	Der Pfeil steht für „reagiert zu“ und trennt die Seite der Edukte von der Seite der Produkte.
Zahlen, zum Beispiel „2“	Die Zahlen stehen für die Anzahl von Teilchen.
tiefgestellte Zahlen, zum Beispiel „ ₂ “	Die tiefgestellten Zahlen stehen für die Anzahl von einzelnen Atomen <i>innerhalb</i> eines Moleküls.

Aufgabe

Um berechnen zu können wieviel Kohlendioxid bei der Verbrennung von einer bestimmten Menge Kohle frei wird, müssen die Greenpeace-Mitarbeiterinnen Sandra und Antje zunächst verstehen, wie eine chemische Reaktionsgleichung funktioniert. Die beiden meinen, dass es dafür wichtig ist auf bestimmte **Regeln bei den Veränderungen zwischen den Einzelteilen der Modelle** zu achten. Die Regel lautet in diesem Fall: **„Bei einer chemischen Reaktion werden Elemente weder erschaffen noch vernichtet.“**

Welche Vermutung können sie mit Hilfe der Reaktionsgleichung aufstellen? **Kreuze an!**

Wenn vor einer vollständigen Reaktion Masse, Volumen, Molekülanzahl und Aggregatzustand der Edukte bestimmt werden, dann müssen die Produkte nach dieser Reaktion insgesamt ...

- ... das gleiche Volumen wie die Edukte haben.
- ... die gleiche Masse wie die Edukte haben.
- ... die gleiche Anzahl Moleküle wie die Edukte haben.
- ... den gleichen Aggregatzustand wie die Edukte haben.

I. Lebenslauf inkl. Publikationen

Persönliche Daten

Name: Marvin Rost
Anschrift: Schloßstraße 63, 14059 Berlin
Geburtsdatum: 02.06.1986
Geburtsort: Berlin
Telefon: 0172 39 211 59
Mail: mailanrost@gmail.com
Familienstand: ledig

Ausbildung/Studium

10/2015–laufend **Promotionsphase des "ProLeA-Promotionsstudiengangs"**, *Humboldt-Universität zu Berlin*.

04/2015–09/2015 **Masterphase des "ProLeA-Promotionsstudiengangs"**, *Humboldt-Universität zu Berlin*.

09/2013–09/2015 **Studium: Master of Education**, *Humboldt-Universität zu Berlin*, Fächer: Chemie und Philosophie/Ethik.
Abschlussnote: 1,5

10/2009–09/2013 **Studium: Bachelor of Science**, *Freie Universität Berlin*, Fächer: Chemie und Ethik.
Abschlussnote: 2,0

06/2009 **Fremdenprüfung: staatl. anerkannter chem.-tech. Assistent**, *OSZ Lise-Meitner*, Berlin.
Notendurchschnitt: 2,7

11/2005–01/2009 **IHK-Ausbildung: Chemielaborant**, *Bayer AG*, Berlin.
Abschlussbewertung: 92% $\hat{=}$ sehr gut

05/2005 **Schulabschluss: Allgemeine Hochschulreife**, *Europäisches Gymnasium Bertha-von-Suttner*, Berlin.
Abschlussnote: 2,4

Berufliche Tätigkeiten

01/2020–laufend **Wissenschaftlicher Mitarbeiter: Organisation und Durchführung von Lehrveranstaltungen**, *Lehrstuhl: Fachdidaktik und Lehr-/Lernforschung Chemie der Humboldt-Universität zu Berlin*, Lehrstuhlinhaber: Prof. Dr. Rüdiger Tiemann.

01/2019–06/2019 **Vertretungslehrer: Chemie und integrierte Naturwissenschaften**, *Europäisches Gymnasium Bertha-von-Suttner (7 WoStd.) und Käthe-Kollwitz-Gymnasium (9 WoStd.)*, Berlin.

10/2017–07/2018 **Wissenschaftlicher Mitarbeiter: Konzeption, Organisation und Durchführung einer Lehrveranstaltung für Studierende des Grundschullehramts für Naturwissenschaften**, *Lehrstuhl: Fachdidaktik und Lehr-/Lernforschung Chemie der Humboldt-Universität zu Berlin*, Lehrstuhlinhaber: Prof. Dr. Rüdiger Tiemann.

06/2013–09/2015 **Studentische Hilfskraft: Lehrveranstaltungsorganisation und Peer-Mentoring von Lehramtsstudierenden zu Studienbeginn**, *Zentrum für Lehrerbildung & Arbeitsbereich Lernpsychologie der Freien Universität Berlin*, Lehrstuhlinhaber: Prof. Dr. Annette Kinder.

01/2011–08/2011 **Studentische Hilfskraft: Chemische Laborarbeiten in der Mikrochipforschung und -entwicklung**, *Technische Universität Berlin/Fraunhofer-Institut für Zuverlässigkeit und Mikrointegration IZM*, Berlin.

I. Lebenslauf inkl. Publikationen

- Auslandsaufenthalte
- 11/2018–01/2019 **Forschungsaufenthalt**, Weizmann-Institut für Naturwissenschaften, Rehovot (Israel).
- 01/2009–05/2009 **Praktikum: Chemielaborant**, Ölraffinerie der Preem AB, Lysekil (Schweden).
- Besuch wissenschaftlicher Tagungen
- Vorträge
- 09/2019 **Empirische Struktur von Modellnutzung im Chemieunterricht**, *GDCP-Jahrestagung*, Wien, mit Prof. Dr. R. Tiemann.
- 08/2019 **Empirical Structure of Using Models in Chemistry Education**, *ESERA-Tagung*, Bologna (Italien), mit Prof. Dr. R. Tiemann.
- 04/2019 **Empirical Evidence for Describing Mental Models in Chemistry Education**, *NARST-Tagung*, Baltimore (MD, USA), mit Prof. Dr. R. Tiemann.
- 10/2018 **Modelle im Chemieunterricht**, 1. Bundeskongress "Jugend präsentiert" an der Berlin-Brandenburgischen Akademie der Wissenschaften, Berlin, mit Prof. Dr. R. Tiemann.
- 09/2017 **Erkenntnisgewinnung mit Modellen im Chemieunterricht**, *GDCP-Jahrestagung*, Regensburg, mit Prof. Dr. R. Tiemann.
- 10/2015 **Modellkompetenz im Kontext Chemieunterricht**, *Nachwuchstagung des IZBF der Humboldt-Universität zu Berlin*, mit Prof. Dr. R. Tiemann.
- Poster
- 02/2019 **Modellnutzung und Erkenntnisgewinnung**, *Tag der Bildungsforschung, IZfB der Universität Duisburg-Essen*, Essen, mit Prof. Dr. R. Tiemann.
- 07/2017 **Scientific Reasoning with Models**, *Tag der Chemie*, Humboldt-Universität zu Berlin, mit Prof. Dr. R. Tiemann.
- 04/2017 **Scientific Reasoning with Models**, *NARST-Tagung*, San Antonio (TX, USA), mit Prof. Dr. R. Tiemann.
- 03/2017 **Modelle im Chemieunterricht nutzen**, *GEBF-Jahrestagung*, Heidelberg, mit Prof. Dr. R. Tiemann.
- 09/2016 **Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht nutzen**, *GDCP-Jahrestagung*, Zürich, mit Prof. Dr. R. Tiemann.
- 09/2015 **Literaturüberblick zur Modellbildung als Mittel der Erkenntnisgewinnung**, *GDCP-Jahrestagung*, Berlin, mit Prof. Dr. R. Tiemann.
- Workshops als Dozent
- 09/2018 **Sprachförderung im Chemieunterricht - Geht das?**, mit der AG Bildungsstandards auf der Jahrestagung der Fachgruppe Chemieunterricht der GDCh, Karlsruhe.
- 09/2017 **Wie lässt sich Erkenntnisgewinnung bei Schüler*innen durch Nutzung von Modellen im Chemieunterricht fördern? Eine offene Werkstatt**, mit der AG Bildungsstandards auf der Jahrestagung der Fachgruppe Chemieunterricht der GDCh, Berlin.
- 09/2016 **Kompetenzorientierter Chemieunterricht durch Förderung von Modellkompetenz - eine offene Werkstatt**, mit der AG Bildungsstandards auf der Jahrestagung der Fachgruppe Chemieunterricht der GDCh, Hannover.
- ohne eigenen Beitrag
- 09/2018 **Netzwerktreffen**, *GDCP-Jahrestagung*, Kiel.
- 03/2016 **Besuch ausgewählter, projektbezogener Vorträge**, *GEBF-Jahrestagung*, Berlin.

- 03/2015 **Besuch von Methodenworkshops & Teambuildingmaßnahme für das "ProLeA-Promotionsprogramm"**, *GEBF-Nachwuchstagung*, Bochum.
- 09/2014 **erster Tagungsbesuch während des Masterstudiums**, *GDCP-Jahrestagung*, Bremen.

Publikationen

Tagungsbände

- Rost, M. & Tiemann, R. (2020): **Empirische Struktur von Modellnutzung im Chemieunterricht**. In S. Habig (Hg.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Wien. Universität Duisburg-Essen. S. 532–535.
- Rost, M. & Tiemann, R. (2018): **Aufgabenentwicklung zur Erhebung der Modellnutzung für den Kompetenzbereich der Erkenntnisgewinnung im Chemieunterricht**. In C. Maurer (Hg.), *Qualitätsvoller Chemie- und Physikunterricht - normative und empirische Dimensionen*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Regensburg 2017. Universität Regensburg. S. 440–443.
- Rost, M. & Tiemann, R. (2017): **Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht. Ein systematischer Überblick ausgewählter empirischer Studien**. In J. Stiller & C. Laschke (Hg.), *Berlin-Brandenburger Beiträge zur Bildungsforschung. Herausforderungen, Befunde und Perspektiven interdisziplinärer Bildungsforschung*. Frankfurt a. M. u. a.: Peter Lang. S. 283–303. [peer-reviewed]
- Rost, M. & Tiemann, R. (2017): **Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht nutzen. Erste Ergebnisse aus einem Ansatz zur Zweckorientierung von Modellen**. In C. Maurer (Hg.), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Zürich 2016. Universität Regensburg. S. 732–735.
- Rost, M. & Tiemann, R. (2016): **Modellbildung als Mittel der Erkenntnisgewinnung – Eine Reviewstudie**. In C. Maurer (Hg.), *Authentizität und Lernen – das Fach in der Fachdidaktik*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Berlin 2015. Universität Regensburg. S. 443–445.

Abschlussarbeiten

- Rost, M.: **Modellbildung als Mittel der Erkenntnisgewinnung im Chemieunterricht. Ein systematischer Überblick ausgewählter empirischer Studien**. *Masterarbeit an der Fachdidaktik und Lehr-/Lernforschung Chemie der Humboldt-Universität zu Berlin*. Erstgutachter: Prof. Dr. Rüdiger Tiemann, Zweitgutachterin: Prof. Dr. Annette Upmeyer zu Belzen. Datum der Abgabe: 04.08.2015.
- Rost, M.: **Chinin in Getränken – Experimentelle Zugänge zur kritischen Auseinandersetzung mit Lebensmittelzusatzstoffen**. *Bachelorarbeit an der Didaktik der Chemie der Freien Universität Berlin*. Erstgutachterin: Prof. Dr. Sabine Steller, Zweitgutachter: Prof. Dr. Klaus Roth. Datum der Abgabe: 07.01.2013.

Eigene Lehrveranstaltungen

Universitäre Lehre

- SoSe 2020 **Materialchemie für Grundschullehrantsstudierende**, *Digital geführtes Seminar zur Exploration chemischer Phänomene in Küche und Badezimmer*, 2 SWS.
- WiSe 2019/20 **Einführung in die fachlichen Grundlagen der Chemie**, *Vorlesung im Bachelorstudiengang für Studierende des Grundschullehrants mit dem Schwerpunkt Naturwissenschaften*, 4 SWS.

I. Lebenslauf inkl. Publikationen

- WiSe 2017/18 **Einführung in die fachlichen Grundlagen der Chemie**, Vorlesung im Bachelorstudiengang für Studierende des Grundschullehramts mit dem Schwerpunkt Naturwissenschaften, 2 von 4 SWS.
- SoSe 2017 **Einführung in die Fachdidaktik Chemie**, Grundlagenseminar im Bachelorstudiengang für Lehramtsstudierende des Fachs Chemie, 2 SWS.
- WiSe 2016/17 **Aufbauseminar der Fachdidaktik Chemie**, Vertiefendes Seminar im Bachelorstudiengang für Lehramtsstudierende des Fachs Chemie, 2 SWS.
- SoSe 2016 **Einführung in die Fachdidaktik Chemie**, Grundlagenseminar im Bachelorstudiengang für Lehramtsstudierende des Fachs Chemie, 2 SWS.
- WiSe 2015/16 **Fachbezogene Vermittlungskompetenzen in der Chemie**, Vertiefendes Seminar im Bachelorstudiengang für Lehramtsstudierende des Fachs Chemie, 2 SWS.
- Betreuung von Masterarbeiten
- 10/2019 **Modellbestandteile als Indikatoren für hypothetisch-deduktives Denken in Chemieaufgaben - eine deskriptive Analyse**, angefertigt von Helena Tran, Erstgutachter: Prof. Dr. Rüdiger Tiemann, Zweitgutachterin: Prof. Dr. Annette Upmeyer zu Belzen.
- 09/2017 **Beitrag zur Konstruktvalidierung der naturwissenschaftlichen Arbeitsweise Beobachten-Vergleichen-Ordnen in den Fächern Chemie und Biologie anhand strukturanalytischer Charakterisierungen von Erhebungsaufgaben**, angefertigt von Annika Meißner, Erstgutachter: Prof. Dr. Rüdiger Tiemann, Zweitgutachterin: Prof. Dr. Annette Upmeyer zu Belzen.
- 06/2017 **Formale Argumentationsrekonstruktion in der Chemiedidaktik - Eine Machbarkeitsstudie zu Transfermöglichkeiten**, angefertigt von Sascha Hartmann, Erstgutachter: Prof. Dr. Rüdiger Tiemann, Zweitgutachterin: Prof. Dr. Jenna Koenen.
- Schulunterricht
2. Halbjahr 2019 **Leistungskurs: Chemie**, Reaktionskinetik, Massenwirkungsgesetz, pH-Wert.
2. Halbjahr 2019 **Profilierungskurs: Chemie**, klassische Trennungsgänge der analytischen Chemie.
2. Halbjahr 2019 **10. Klasse: Chemie**, Grundlagen der organischen Chemie.
2. Halbjahr 2019 **7. Klasse: Integrierte Naturwissenschaften**, hypothesengeleitetes Experimentieren, Teilchenmodell, Sexualkunde.
- Lehrerfortbildungen
- 02/2020–laufend **Organisation und Durchführung der Module "Naturwissenschaftliche Arbeitsweisen", "Von den Sinnen zum Messen", "Pflanzen, Tiere, Lebensräume", "Stoffe und Stoffeigenschaften", "Sexualkunde", "Arbeitssicherheit und Gefährdungsbeurteilungen" & "Selbst- und Unterrichtsreflexion im Kontext naturwissenschaftlichen Unterrichts"**, 8. Durchgang der Fortbildung zur Einführung des Rahmenlehrplans "Naturwissenschaften in der 5./6. Klasse" durch das Bildungsministerium des Landes Brandenburg über den W.i.B. e. V. (An-Institut der Universität Potsdam), 8 Std. & ca. 25 TeilnehmerInnen pro Modul, jeweils als 1 von 2 DozentInnen.
- 08/2019–11/2019 **Organisation und Durchführung der Module "Naturwissenschaftliche Arbeitsweisen", "Pflanzen, Tiere, Lebensräume", "Stoffe und Stoffeigenschaften", "Sexualkunde", "Arbeitssicherheit und Gefährdungsbeurteilungen" & "Selbst- und Unterrichtsreflexion im Kontext naturwissenschaftlichen Unterrichts"**, 7. Durchgang der Fortbildung zur Einführung des Rahmenlehrplans "Naturwissenschaften in der 5./6. Klasse" durch das Bildungsministerium des Landes Brandenburg über den W.i.B. e. V. (An-Institut der Universität Potsdam), 8 Std. & ca. 25 TeilnehmerInnen pro Modul, jeweils als 1 von 2 DozentInnen.

- 03/2019–05/2019 **Organisation und Durchführung der Module "Von den Sinnen zum Messen", "Die Sonne als Energiequelle", "Pflanzen, Tiere, Lebensräume", "Stoffe und Stoffeigenschaften", "Sexualkunde", "Arbeitssicherheit und Gefährdungsbeurteilungen" & "Selbst- und Unterrichtsreflexion im Kontext naturwissenschaftlichen Unterrichts"**, 6. Durchgang der Fortbildung zur Einführung des Rahmenlehrplans "Naturwissenschaften in der 5./6. Klasse" durch das Bildungsministerium des Landes Brandenburg über den W.i.B. e. V. (An-Institut der Universität Potsdam), 8 Std. & ca. 25 TeilnehmerInnen pro Modul, jeweils als 1 von 2 DozentInnen.
- 03/2018–04/2019 **Organisation und Durchführung der Module "Von den Sinnen zum Messen", "Die Sonne als Energiequelle", "Kompetenzen im naturwissenschaftlichen Unterricht", "Arbeitssicherheit und Gefährdungsbeurteilungen" & "Selbst- und Unterrichtsreflexion im Kontext naturwissenschaftlichen Unterrichts"**, 5. Durchgang der Fortbildung zur Einführung des Rahmenlehrplans "Naturwissenschaften in der 5./6. Klasse" durch das Bildungsministerium des Landes Brandenburg über den W.i.B. e. V. (An-Institut der Universität Potsdam), 9 Std. & ca. 100 TeilnehmerInnen pro Modul, jeweils als 1 von 2 DozentInnen.
- 11/2017–01/2019 **Organisation und Durchführung der Module "Von den Sinnen zum Messen", "Die Sonne als Energiequelle", "Kompetenzen im naturwissenschaftlichen Unterricht", "Arbeitssicherheit und Gefährdungsbeurteilungen" & "Selbst- und Unterrichtsreflexion im Kontext naturwissenschaftlichen Unterrichts"**, 4. Durchgang der Fortbildung zur Einführung des Rahmenlehrplans "Naturwissenschaften in der 5./6. Klasse" durch das Bildungsministerium des Landes Brandenburg über den W.i.B. e. V. (An-Institut der Universität Potsdam), 9 Std. & ca. 100 TeilnehmerInnen pro Modul, jeweils als 1 von 2 DozentInnen.
- 06/2018 **Konzeption, Organisation und Durchführung von "Chemie für Anfänger"**, Fortbildung der Fachdidaktik und Lehr-/Lernforschung Chemie der Humboldt-Universität zu Berlin für LehrerInnen an einer Schule mit sonderpädagogischem Förderschwerpunkt, 7 Std. & 6 TeilnehmerInnen.
- 07/2017–05/2018 **Organisation und Durchführung der Module "Von den Sinnen zum Messen", "Die Sonne als Energiequelle", "Arbeitssicherheit und Gefährdungsbeurteilungen" & "Selbst- und Unterrichtsreflexion im Kontext naturwissenschaftlichen Unterrichts"**, 3. Durchgang der Fortbildung zur Einführung des Rahmenlehrplans "Naturwissenschaften in der 5./6. Klasse" durch das Bildungsministerium des Landes Brandenburg über den W.i.B. e. V. (An-Institut der Universität Potsdam), 9 Std. & ca. 100 TeilnehmerInnen pro Modul, jeweils als 1 von 2 DozentInnen.
- 04/2017–05/2017 **Organisation und Durchführung der Module "Von den Sinnen zum Messen" & "Die Sonne als Energiequelle"**, 2. Durchgang der Fortbildung zur Einführung des Rahmenlehrplans "Naturwissenschaften in der 5./6. Klasse" durch das Bildungsministerium des Landes Brandenburg über den W.i.B. e. V. (An-Institut der Universität Potsdam), 9 Std. & ca. 100 TeilnehmerInnen pro Modul, jeweils als 1 von 2 DozentInnen.
- 11/2016–02/2017 **Konzeption, Organisation und Durchführung der Module "Von den Sinnen zum Messen" & "Die Sonne als Energiequelle"**, 1. Durchgang der Fortbildung zur Einführung des Rahmenlehrplans "Naturwissenschaften in der 5./6. Klasse" durch das Bildungsministerium des Landes Brandenburg über den W.i.B. e. V. (An-Institut der Universität Potsdam), 9 Std. & ca. 100 TeilnehmerInnen pro Modul, jeweils als 1 von 2 DozentInnen.

I. Lebenslauf inkl. Publikationen

Einladungen als Dozent

- 10/2019 **Sprachbildung im naturwissenschaftlichen Unterricht**, *Lehrerfortbildung durch das pädagogische Landesinstitut Rheinland-Pfalz mit der AG Biologiedidaktik der Universität Koblenz-Landau*, Landau, mit Prof. Dr. S. Nitz.
- 09/2018 **Das Parfum steht Ihnen! Warum auch Moleküle gut passen müssen**, *Keynote auf dem "Tag der Würdigung" des Käthe-Kollwitz-Gymnasiums*, Berlin.
- 01/2018 **Von der Berufsausbildung zur Promotion - Optionen nach dem Abitur**, *Vortrag auf dem "Career Day" des Europäischen Gymnasiums Bertha-von-Suttner*, Berlin.

Drittmittelinwerbung

- 08/2019 **Reisemittel zur Finanzierung der ESERA-Tagung in Bologna (Italien)**, *Deutscher Akademischer Austauschdienst*.
- 10/2015–09/2017 **Stipendiat des "ProLeA-Promotionsprogramms"**, *Humboldt-Universität zu Berlin*.

Gutachtertätigkeiten

- 10/2018 **Begutachtung von zwei Einreichungen für die NARST-Tagung 2019**.

Sprachkenntnisse

- Englisch fließend
Französisch Grundkenntnisse
Schwedisch Grundkenntnisse

Mitgliedschaften

- Gesellschaft für Didaktik der Chemie und Physik (*GDCP*)
- Gesellschaft Deutscher Chemiker (*GDCh*), Fachgruppe Chemieunterricht

Sonstiges

- Träger des silbernen Feuerwehr- und Katastrophenschutz-Ehrenzeichens am Bande des Landes Berlin für langjähriges Engagement im ehrenamtlichen Wasserrettungsdienst der Deutschen Lebensrettungsgesellschaft e. V. in Berlin.

J. Eigenständigkeitserklärung

Ich erkläre, dass ich die Dissertation selbständig und nur unter Verwendung der von mir gemäß § 7 Abs. 3 der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät, veröffentlicht im Amtlichen Mitteilungsblatt der Humboldt-Universität zu Berlin Nr. 42/2018 am 11.07.2018 angegebenen Hilfsmittel angefertigt habe.

Marvin Rost
Berlin,

K. Danksagung

Ich habe diese Arbeit ganz allein geschrieben und bin vor einigen Jahren einmal darauf angesprochen worden, dass Danksagungen von Qualifikationsarbeiten gelegentlich so ausschweifend sind, dass sie die unterschriebene Eigenständigkeitserklärung ad absurdum führen würden. Das kann ich so nicht nachvollziehen: Die Unterstützung durch eine Vielzahl von Menschen in meinem Umfeld war für mich eine notwendige Bedingung, um in der Lage zu sein diese Arbeit anfangen und zum Abschluss bringen zu können.

Meinem Doktorvater, Prof. Dr. Rüdiger Tiemann, sei hier als erstes gedankt. Er war nicht nur offen für meine Ambitionen, die Verbindungen zwischen Empirie und Erkenntnistheorie auszuloten – er hat mich dazu ermutigt und meine eigenen Ideen von Beginn an gefördert. Darüber hinaus hat er spielerische Herausforderungen gesehen und kreative Herangehensweisen gesucht, während ich noch versuchte analytische Probleme zu wälzen. Das wird meine weiteren Vorhaben prägen.

Prof. Dr. Jenna Koenen danke ich für eine unzählbare Menge konstruktiven Inputs, zum Beispiel bei Proposals für Tagungen und viele schöne Pausengespräche. Eine Randbemerkung in einem dieser Gespräche gab die Initialzündung für die statistische Struktur von Modellnutzung in dieser Arbeit und ich bedauere, dass ich die ursprüngliche Papier- und Bleistift-Skizze nicht mehr finde!

Der ganze Arbeitskreis hat mich mit der warmen und lustigen Atmosphäre auch in schwierigen Momenten getragen. Namentlich möchte ich Renate Glindkamp, Dr. Petra Esperling und Simon Schäfer danken. Renate hat mir sehr bei der Datenübertragung geholfen und bei jeder Gelegenheit schneller ein praktisches Laborproblem gelöst als ich es vollständig zu formulieren in der Lage war. Petra hat mir bei jeglichen Verwaltungsvorgängen immer wieder geduldig geholfen und damit dazu beigetragen, dass ich mich auf meine wissenschaftliche Arbeit konzentrieren konnte. Das empfand ich nicht als Selbstverständlichkeit. Obwohl sie nicht zum Arbeitskreis gehört, es aber gerade passt, danke ich an dieser Stelle auch Nora Butter, die mich bei der Vor- und Nachbereitung meines Auslandsaufenthalts sehr unterstützt hat. Ohne Sie wäre die Planung dieser schönen Zeit deutlich schwieriger geworden. Simon Schäfer möchte ich dankend erwähnen, weil er bei der Kombinatorik des Testheftdesigns weit über eine Beratung vom Spielfeldrand hinaus gegangen ist.

An alle teilnehmenden SchülerInnen geht ein anonymer Dank. Auch ohne die zahlreichen LehrerInnen, die bereit waren mir ihre wertvolle Unterrichtszeit zur Verfügung zu stellen, wäre das Projekt nicht über die Theoriephase hinausgegangen. Drei von ihnen möchte besonders erwähnen. Alexander Lotz hat mir durch sein Netzwerk Türen geöffnet und das werde ich nicht vergessen. Er gehört zur AG Bildungsstandards der Gesellschaft Deutscher Chemiker, in deren Sitzungen ich VeteranInnen der Schulwelt gegenüber saß. Ich danke jedem einzelnen für die von Wertschätzung geprägten Gesprä-

K. Danksagung

che und die Erdung meines Projekts durch die kritische Prüfung an der Dienst erfahrung gestandener Lehrkräfte. In diese Kategorie fallen auch Nicola Leu und Steffen Sladek, die mich bei der Suche nach teilnehmenden Schulen ebenfalls stark unterstützt haben. Unsere zahlreichen Sit-Ins oszillierten zwischen hochgestochener Kritik am Wesen des Bildungssystems und einem Humor, der mir Pausen ermöglichte. Das war wichtig.

Vier Studierende möchte ich ebenfalls erwähnen. Annika Meißner hat bei der Erhebung und Administration der Daten geholfen. Ihre Masterarbeit ermöglichte eine systematische Überarbeitung der Aufgabenstämme der Studie, was die Qualität deutlich erhöhte. Ohne Svenja Ahrens wäre die Stichprobe deutlich kleiner ausgefallen und ihre Unterstützung in der Lehre hat eine ganze Vorlesung besser gemacht. Sascha Hartmanns Masterarbeit war der Auftakt für die Überlegungen zur Verknüpfung von Modellnutzung und Argumentationstheorie und ich bin sehr froh, dass er sich auf diese explorative Arbeit eingelassen hat. Helena Tran hat im Rahmen ihrer Masterarbeit ganz eigenständig ein paar ausgewählte Aufgaben an SchülerInnen ausgegeben. Ihr Fleiß ermöglicht mir unmittelbare empirische Anschlussforschung und das ist unbezahlbar.

Den Wegbegleitern aus dem ProLeA-Programm der Humboldt-Universität danke ich für jedes Seminar, jede Diskussion, jedes Gespräch und jedes Stück selbstgebackene Pizza. Jede einzelne von ihnen hat dazu beigetragen mich mit der Universität zu identifizieren und Teil an etwas zu haben. Den treibenden Akteuren hinter dem Programm danke ich ebenfalls sehr. Im Rahmen des Programmstipendiums war eine entscheidende Zeit des Projekts finanziert und mir wurden wichtige Tagungen und Weiterbildungsmaßnahmen ermöglicht, ohne die die Arbeit eine andere – vermutlich keine bessere – geworden wäre.

Meiner Familie und meinen engen Freunden kann ich nicht namentlich danken. Ich wüsste nicht wie ich anfangen und wie ich aufhören sollte. Sie alle haben sich immer mit mir gefreut wenn ich etwas geschafft hatte und sie alle haben ihre Wertschätzung und Liebe für mich völlig unabhängig von akademischen Erfolgen zum Ausdruck gebracht. Ohne wäre es nicht gegangen.

Bisher erschienene Bände der Reihe „*Studien zum Physik- und Chemielernen*“

ISSN 1614-8967 (vormals *Studien zum Physiklernen* ISSN 1435-5280)

- 1 Helmut Fischler, Jochen Peuckert (Hrsg.): Concept Mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie
ISBN 978-3-89722-256-4 40.50 EUR
- 2 Anja Schoster: Bedeutungsentwicklungsprozesse beim Lösen algorithmischer Physikaufgaben. *Eine Fallstudie zu Lernprozessen von Schülern im Physiknachhilfeunterricht während der Bearbeitung algorithmischer Physikaufgaben*
ISBN 978-3-89722-045-4 40.50 EUR
- 3 Claudia von Aufschnaiter: Bedeutungsentwicklungen, Interaktionen und situatives Erleben beim Bearbeiten physikalischer Aufgaben
ISBN 978-3-89722-143-7 40.50 EUR
- 4 Susanne Haeberlen: Lernprozesse im Unterricht mit Wasserstromkreisen. *Eine Fallstudie in der Sekundarstufe I*
ISBN 978-3-89722-172-7 40.50 EUR
- 5 Kerstin Haller: Über den Zusammenhang von Handlungen und Zielen. *Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*
ISBN 978-3-89722-242-7 40.50 EUR
- 6 Michaela Horstendahl: Motivationale Orientierungen im Physikunterricht
ISBN 978-3-89722-227-4 50.00 EUR
- 7 Stefan Deylitz: Lernergebnisse in der Quanten-Atomphysik. *Evaluation des Bremer Unterrichtskonzepts*
ISBN 978-3-89722-291-5 40.50 EUR
- 8 Lorenz Hucke: Handlungsregulation und Wissenserwerb in traditionellen und computergestützten Experimenten des physikalischen Praktikums
ISBN 978-3-89722-316-5 50.00 EUR
- 9 Heike Theyßen: Ein Physikpraktikum für Studierende der Medizin. *Darstellung der Entwicklung und Evaluation eines adressatenspezifischen Praktikums nach dem Modell der Didaktischen Rekonstruktion*
ISBN 978-3-89722-334-9 40.50 EUR
- 10 Annette Schick: Der Einfluß von Interesse und anderen selbstbezogenen Kognitionen auf Handlungen im Physikunterricht. *Fallstudien zu Interessenhandlungen im Physikunterricht*
ISBN 978-3-89722-380-6 40.50 EUR
- 11 Roland Berger: Moderne bildgebende Verfahren der medizinischen Diagnostik. *Ein Weg zu interessanterem Physikunterricht*
ISBN 978-3-89722-445-2 40.50 EUR

- 12 Johannes Werner: Vom Licht zum Atom. *Ein Unterrichtskonzept zur Quantenphysik unter Nutzung des Zeigermodells*
ISBN 978-3-89722-471-1 40.50 EUR
- 13 Florian Sander: Verbindung von Theorie und Experiment im physikalischen Praktikum. *Eine empirische Untersuchung zum handlungsbezogenen Vorverständnis und dem Einsatz grafikorientierter Modellbildung im Praktikum*
ISBN 978-3-89722-482-7 40.50 EUR
- 14 Jörn Gerdes: Der Begriff der physikalischen Kompetenz. *Zur Validierung eines Konstruktes*
ISBN 978-3-89722-510-7 40.50 EUR
- 15 Malte Meyer-Arndt: Interaktionen im Physikpraktikum zwischen Studierenden und Betreuern. *Feldstudie zu Bedeutungsentwicklungsprozessen im physikalischen Praktikum*
ISBN 978-3-89722-541-1 40.50 EUR
- 16 Dietmar Höttecke: Die Natur der Naturwissenschaften historisch verstehen. *Fachdidaktische und wissenschaftshistorische Untersuchungen*
ISBN 978-3-89722-607-4 40.50 EUR
- 17 Gil Gabriel Mavanga: Entwicklung und Evaluation eines experimentell- und phänomenorientierten Optikcurriculums. *Untersuchung zu Schülervorstellungen in der Sekundarstufe I in Mosambik und Deutschland*
ISBN 978-3-89722-721-7 40.50 EUR
- 18 Meike Ute Zastrow: Interaktive Experimentieranleitungen. *Entwicklung und Evaluation eines Konzeptes zur Vorbereitung auf das Experimentieren mit Messgeräten im Physikalischen Praktikum*
ISBN 978-3-89722-802-3 40.50 EUR
- 19 Gunnar Friege: Wissen und Problemlösen. *Eine empirische Untersuchung des wissenszentrierten Problemlösens im Gebiet der Elektrizitätslehre auf der Grundlage des Experten-Novizen-Vergleichs*
ISBN 978-3-89722-809-2 40.50 EUR
- 20 Erich Starauschek: Physikunterricht nach dem Karlsruher Physikkurs. *Ergebnisse einer Evaluationsstudie*
ISBN 978-3-89722-823-8 40.50 EUR
- 21 Roland Paatz: Charakteristika analogiebasierten Denkens. *Vergleich von Lernprozessen in Basis- und Zielbereich*
ISBN 978-3-89722-944-0 40.50 EUR
- 22 Silke Mikelskis-Seifert: Die Entwicklung von Metakzepten zur Teilchenvorstellung bei Schülern. *Untersuchung eines Unterrichts über Modelle mithilfe eines Systems multipler Repräsentationsebenen*
ISBN 978-3-8325-0013-9 40.50 EUR
- 23 Brunhild Landwehr: Distanzen von Lehrkräften und Studierenden des Sachunterrichts zur Physik. *Eine qualitativ-empirische Studie zu den Ursachen*
ISBN 978-3-8325-0044-3 40.50 EUR

- 24 Lydia Murmann: Physiklernen zu Licht, Schatten und Sehen. *Eine phänomenografische Untersuchung in der Primarstufe*
ISBN 978-3-8325-0060-3 40.50 EUR
- 25 Thorsten Bell: Strukturprinzipien der Selbstregulation. *Komplexe Systeme, Elementarisierungen und Lernprozessstudien für den Unterricht der Sekundarstufe II*
ISBN 978-3-8325-0134-1 40.50 EUR
- 26 Rainer Müller: Quantenphysik in der Schule
ISBN 978-3-8325-0186-0 40.50 EUR
- 27 Jutta Roth: Bedeutungsentwicklungsprozesse von Physikerinnen und Physikern in den Dimensionen Komplexität, Zeit und Inhalt
ISBN 978-3-8325-0183-9 40.50 EUR
- 28 Andreas Saniter: Spezifika der Verhaltensmuster fortgeschrittener Studierender der Physik
ISBN 978-3-8325-0292-8 40.50 EUR
- 29 Thomas Weber: Kumulatives Lernen im Physikunterricht. *Eine vergleichende Untersuchung in Unterrichtsgängen zur geometrischen Optik*
ISBN 978-3-8325-0316-1 40.50 EUR
- 30 Markus Rehm: Über die Chancen und Grenzen moralischer Erziehung im naturwissenschaftlichen Unterricht
ISBN 978-3-8325-0368-0 40.50 EUR
- 31 Marion Budde: Lernwirkungen in der Quanten-Atom-Physik. *Fallstudien über Resonanzen zwischen Lernangeboten und SchülerInnen-Vorstellungen*
ISBN 978-3-8325-0483-0 40.50 EUR
- 32 Thomas Reyer: Oberflächenmerkmale und Tiefenstrukturen im Unterricht. *Exemplarische Analysen im Physikunterricht der gymnasialen Sekundarstufe*
ISBN 978-3-8325-0488-5 40.50 EUR
- 33 Christoph Thomas Müller: Subjektive Theorien und handlungsleitende Kognitionen von Lehrern als Determinanten schulischer Lehr-Lern-Prozesse im Physikunterricht
ISBN 978-3-8325-0543-1 40.50 EUR
- 34 Gabriela Jonas-Ahrend: Physiklehrvorstellungen zum Experiment im Physikunterricht
ISBN 978-3-8325-0576-9 40.50 EUR
- 35 Dimitrios Stavrou: Das Zusammenspiel von Zufall und Gesetzmäßigkeiten in der nicht-linearen Dynamik. *Didaktische Analyse und Lernprozesse*
ISBN 978-3-8325-0609-4 40.50 EUR
- 36 Katrin Engeln: Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken
ISBN 978-3-8325-0689-6 40.50 EUR
- 37 Susann Hartmann: Erklärungsvielfalt
ISBN 978-3-8325-0730-5 40.50 EUR

- 38 Knut Neumann: Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker
ISBN 978-3-8325-0762-6 40.50 EUR
- 39 Michael Späth: Kontextbedingungen für Physikunterricht an der Hauptschule. *Möglichkeiten und Ansatzpunkte für einen fachübergreifenden, handlungsorientierten und berufsorientierten Unterricht*
ISBN 978-3-8325-0827-2 40.50 EUR
- 40 Jörg Hirsch: Interesse, Handlungen und situatives Erleben von Schülerinnen und Schülern beim Bearbeiten physikalischer Aufgaben
ISBN 978-3-8325-0875-3 40.50 EUR
- 41 Monika Hüther: Evaluation einer hypermedialen Lernumgebung zum Thema Gasgesetze. *Eine Studie im Rahmen des Physikpraktikums für Studierende der Medizin*
ISBN 978-3-8325-0911-8 40.50 EUR
- 42 Maike Tesch: Das Experiment im Physikunterricht. *Didaktische Konzepte und Ergebnisse einer Videostudie*
ISBN 978-3-8325-0975-0 40.50 EUR
- 43 Nina Nicolai: Skriptgeleitete Eltern-Kind-Interaktion bei Chemiehausaufgaben. *Eine Evaluationsstudie im Themenbereich Säure-Base*
ISBN 978-3-8325-1013-8 40.50 EUR
- 44 Antje Leisner: Entwicklung von Modellkompetenz im Physikunterricht
ISBN 978-3-8325-1020-6 40.50 EUR
- 45 Stefan Rumann: Evaluation einer Interventionsstudie zur Säure-Base-Thematik
ISBN 978-3-8325-1027-5 40.50 EUR
- 46 Thomas Wilhelm: Konzeption und Evaluation eines Kinematik/Dynamik-Lehrgangs zur Veränderung von Schülervorstellungen mit Hilfe dynamisch ikonischer Repräsentationen und graphischer Modellbildung – mit CD-ROM
ISBN 978-3-8325-1046-6 45.50 EUR
- 47 Andrea Maier-Richter: Computerunterstütztes Lernen mit Lösungsbeispielen in der Chemie. *Eine Evaluationsstudie im Themenbereich Löslichkeit*
ISBN 978-3-8325-1046-6 40.50 EUR
- 48 Jochen Peuckert: Stabilität und Ausprägung kognitiver Strukturen zum Atombegriff
ISBN 978-3-8325-1104-3 40.50 EUR
- 49 Maik Walpuski: Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback
ISBN 978-3-8325-1184-5 40.50 EUR
- 50 Helmut Fischler, Christiane S. Reiners (Hrsg.): Die Teilchenstruktur der Materie im Physik- und Chemieunterricht
ISBN 978-3-8325-1225-5 34.90 EUR
- 51 Claudia Eysel: Interdisziplinäres Lehren und Lernen in der Lehrerbildung. *Eine empirische Studie zum Kompetenzerwerb in einer komplexen Lernumgebung*
ISBN 978-3-8325-1238-5 40.50 EUR

- 52 Johannes Günther: Lehrerfortbildung über die Natur der Naturwissenschaften. *Studien über das Wissenschaftsverständnis von Grundschullehrkräften*
ISBN 978-3-8325-1287-3 40.50 EUR
- 53 Christoph Neugebauer: Lernen mit Simulationen und der Einfluss auf das Problemlösen in der Physik
ISBN 978-3-8325-1300-9 40.50 EUR
- 54 Andreas Schnirch: Gendergerechte Interessen- und Motivationsförderung im Kontext naturwissenschaftlicher Grundbildung. *Konzeption, Entwicklung und Evaluation einer multimedial unterstützten Lernumgebung*
ISBN 978-3-8325-1334-4 40.50 EUR
- 55 Hilde Köster: Freies Explorieren und Experimentieren. *Eine Untersuchung zur selbstbestimmten Gewinnung von Erfahrungen mit physikalischen Phänomenen im Sachunterricht*
ISBN 978-3-8325-1348-1 40.50 EUR
- 56 Eva Heran-Dörr: Entwicklung und Evaluation einer Lehrerfortbildung zur Förderung der physikdidaktischen Kompetenz von Sachunterrichtslehrkräften
ISBN 978-3-8325-1377-1 40.50 EUR
- 57 Agnes Szabone Varnai: Unterstützung des Problemlösens in Physik durch den Einsatz von Simulationen und die Vorgabe eines strukturierten Kooperationsformats
ISBN 978-3-8325-1403-7 40.50 EUR
- 58 Johannes Rethfeld: Aufgabenbasierte Lernprozesse in selbstorganisationsoffenem Unterricht der Sekundarstufe I zum Themengebiet ELEKTROSTATIK. *Eine Feldstudie in vier 10. Klassen zu einer kartenbasierten Lernumgebung mit Aufgaben aus der Elektrostatik*
ISBN 978-3-8325-1416-7 40.50 EUR
- 59 Christian Henke: Experimentell-naturwissenschaftliche Arbeitsweisen in der Oberstufe. *Untersuchung am Beispiel des HIGHSEA-Projekts in Bremerhaven*
ISBN 978-3-8325-1515-7 40.50 EUR
- 60 Lutz Kasper: Diskursiv-narrative Elemente für den Physikunterricht. *Entwicklung und Evaluation einer multimedialen Lernumgebung zum Erdmagnetismus*
ISBN 978-3-8325-1537-9 40.50 EUR
- 61 Thorid Rabe: Textgestaltung und Aufforderung zu Selbsterklärungen beim Physiklernen mit Multimedia
ISBN 978-3-8325-1539-3 40.50 EUR
- 62 Ina Glemnitz: Vertikale Vernetzung im Chemieunterricht. *Ein Vergleich von traditionellem Unterricht mit Unterricht nach Chemie im Kontext*
ISBN 978-3-8325-1628-4 40.50 EUR
- 63 Erik Einhaus: Schülerkompetenzen im Bereich Wärmelehre. *Entwicklung eines Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells fachbezogener Kompetenzen*
ISBN 978-3-8325-1630-7 40.50 EUR

- 64 Jasmin Neuroth: Concept Mapping als Lernstrategie. *Eine Interventionsstudie zum Chemielernen aus Texten*
ISBN 978-3-8325-1659-8 40.50 EUR
- 65 Hans Gerd Hegeler-Burkhart: Zur Kommunikation von Hauptschülerinnen und Hauptschülern in einem handlungsorientierten und fächerübergreifenden Unterricht mit physikalischen und technischen Inhalten
ISBN 978-3-8325-1667-3 40.50 EUR
- 66 Karsten Rincke: Sprachentwicklung und Fachlernen im Mechanikunterricht. *Sprache und Kommunikation bei der Einführung in den Kraftbegriff*
ISBN 978-3-8325-1699-4 40.50 EUR
- 67 Nina Strehle: Das Ion im Chemieunterricht. *Alternative Schülervorstellungen und curriculare Konsequenzen*
ISBN 978-3-8325-1710-6 40.50 EUR
- 68 Martin Hopf: Problemorientierte Schülerexperimente
ISBN 978-3-8325-1711-3 40.50 EUR
- 69 Anne Beerenwinkel: Fostering conceptual change in chemistry classes using expository texts
ISBN 978-3-8325-1721-2 40.50 EUR
- 70 Roland Berger: Das Gruppenpuzzle im Physikunterricht der Sekundarstufe II. *Eine empirische Untersuchung auf der Grundlage der Selbstbestimmungstheorie der Motivation*
ISBN 978-3-8325-1732-8 40.50 EUR
- 71 Giuseppe Colicchia: Physikunterricht im Kontext von Medizin und Biologie. *Entwicklung und Erprobung von Unterrichtseinheiten*
ISBN 978-3-8325-1746-5 40.50 EUR
- 72 Sandra Winheller: Geschlechtsspezifische Auswirkungen der Lehrer-Schüler-Interaktion im Chemieanfangsunterricht
ISBN 978-3-8325-1757-1 40.50 EUR
- 73 Isabel Wahser: Training von naturwissenschaftlichen Arbeitsweisen zur Unterstützung experimenteller Kleingruppenarbeit im Fach Chemie
ISBN 978-3-8325-1815-8 40.50 EUR
- 74 Claus Brell: Lernmedien und Lernerfolg - reale und virtuelle Materialien im Physikunterricht. *Empirische Untersuchungen in achten Klassen an Gymnasien (Laborstudie) zum Computereinsatz mit Simulation und IBE*
ISBN 978-3-8325-1829-5 40.50 EUR
- 75 Rainer Wackermann: Überprüfung der Wirksamkeit eines Basismodell-Trainings für Physiklehrer
ISBN 978-3-8325-1882-0 40.50 EUR
- 76 Oliver Tepner: Effektivität von Aufgaben im Chemieunterricht der Sekundarstufe I
ISBN 978-3-8325-1919-3 40.50 EUR

- 77 Claudia Geyer: Museums- und Science-Center-Besuche im naturwissenschaftlichen Unterricht aus einer motivationalen Perspektive. *Die Sicht von Lehrkräften und Schülerinnen und Schülern*
ISBN 978-3-8325-1922-3 40.50 EUR
- 78 Tobias Leonhard: Professionalisierung in der Lehrerbildung. *Eine explorative Studie zur Entwicklung professioneller Kompetenzen in der Lehrererstausbildung*
ISBN 978-3-8325-1924-7 40.50 EUR
- 79 Alexander Kauertz: Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben
ISBN 978-3-8325-1925-4 40.50 EUR
- 80 Regina Hübinger: Schüler auf Weltreise. *Entwicklung und Evaluation von Lehr-/Lernmaterialien zur Förderung experimentell-naturwissenschaftlicher Kompetenzen für die Jahrgangsstufen 5 und 6*
ISBN 978-3-8325-1932-2 40.50 EUR
- 81 Christine Waltner: Physik lernen im Deutschen Museum
ISBN 978-3-8325-1933-9 40.50 EUR
- 82 Torsten Fischer: Handlungsmuster von Physiklehrkräften beim Einsatz neuer Medien. *Fallstudien zur Unterrichtspraxis*
ISBN 978-3-8325-1948-3 42.00 EUR
- 83 Corinna Kieren: Chemiehausaufgaben in der Sekundarstufe I des Gymnasiums. *Fragebogenerhebung zur gegenwärtigen Praxis und Entwicklung eines optimierten Hausaufgabendesigns im Themenbereich Säure-Base*
978-3-8325-1975-9 37.00 EUR
- 84 Marco Thiele: Modelle der Thermohalinen Zirkulation im Unterricht. *Eine empirische Studie zur Förderung des Modellverständnisses*
ISBN 978-3-8325-1982-7 40.50 EUR
- 85 Bernd Zinn: Physik lernen, um Physik zu lehren. *Eine Möglichkeit für interessanteren Physikunterricht*
ISBN 978-3-8325-1995-7 39.50 EUR
- 86 Esther Klaes: Außerschulische Lernorte im naturwissenschaftlichen Unterricht. *Die Perspektive der Lehrkraft*
ISBN 978-3-8325-2006-9 43.00 EUR
- 87 Marita Schmidt: Kompetenzmodellierung und -diagnostik im Themengebiet Energie der Sekundarstufe I. *Entwicklung und Erprobung eines Testinventars*
ISBN 978-3-8325-2024-3 37.00 EUR
- 88 Gudrun Franke-Braun: Aufgaben mit gestuften Lernhilfen. *Ein Aufgabenformat zur Förderung der sachbezogenen Kommunikation und Lernleistung für den naturwissenschaftlichen Unterricht*
ISBN 978-3-8325-2026-7 38.00 EUR
- 89 Silke Klos: Kompetenzförderung im naturwissenschaftlichen Anfangsunterricht. *Der Einfluss eines integrierten Unterrichtskonzepts*
ISBN 978-3-8325-2133-2 37.00 EUR

- 90 Ulrike Elisabeth Burkard: Quantenphysik in der Schule. *Bestandsaufnahme, Perspektiven und Weiterentwicklungsmöglichkeiten durch die Implementation eines Medienservers*
ISBN 978-3-8325-2215-5 43.00 EUR
- 91 Ulrike Gromadecki: Argumente in physikalischen Kontexten. *Welche Geltungsgründe halten Physikanfänger für überzeugend?*
ISBN 978-3-8325-2250-6 41.50 EUR
- 92 Jürgen Bruns: Auf dem Weg zur Förderung naturwissenschaftsspezifischer Vorstellungen von zukünftigen Chemie-Lehrenden
ISBN 978-3-8325-2257-5 43.50 EUR
- 93 Cornelius Marsch: Räumliche Atomvorstellung. *Entwicklung und Erprobung eines Unterrichtskonzeptes mit Hilfe des Computers*
ISBN 978-3-8325-2293-3 82.50 EUR
- 94 Maja Brückmann: Sachstrukturen im Physikunterricht. *Ergebnisse einer Videostudie*
ISBN 978-3-8325-2272-8 39.50 EUR
- 95 Sabine Fechner: Effects of Context-oriented Learning on Student Interest and Achievement in Chemistry Education
ISBN 978-3-8325-2343-5 36.50 EUR
- 96 Clemens Nagel: eLearning im Physikalischen Anfängerpraktikum
ISBN 978-3-8325-2355-8 39.50 EUR
- 97 Josef Riese: Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften
ISBN 978-3-8325-2376-3 39.00 EUR
- 98 Sascha Bernholt: Kompetenzmodellierung in der Chemie. *Theoretische und empirische Reflexion am Beispiel des Modells hierarchischer Komplexität*
ISBN 978-3-8325-2447-0 40.00 EUR
- 99 Holger Christoph Stawitz: Auswirkung unterschiedlicher Aufgabenprofile auf die Schülerleistung. *Vergleich von Naturwissenschafts- und Problemlöseaufgaben der PISA 2003-Studie*
ISBN 978-3-8325-2451-7 37.50 EUR
- 100 Hans Ernst Fischer, Elke Sumfleth (Hrsg.): nwu-essen – 10 Jahre Essener Forschung zum naturwissenschaftlichen Unterricht
ISBN 978-3-8325-3331-1 40.00 EUR
- 101 Hendrik Härtig: Sachstrukturen von Physikschulbüchern als Grundlage zur Bestimmung der Inhaltsvalidität eines Tests
ISBN 978-3-8325-2512-5 34.00 EUR
- 102 Thomas Grüß-Niehaus: Zum Verständnis des Löslichkeitskonzeptes im Chemieunterricht. *Der Effekt von Methoden progressiver und kollaborativer Reflexion*
ISBN 978-3-8325-2537-8 40.50 EUR

- 103 Patrick Bronner: Quantenoptische Experimente als Grundlage eines Curriculums zur Quantenphysik des Photons
ISBN 978-3-8325-2540-8 36.00 EUR
- 104 Adrian Voßkühler: Blickbewegungsmessung an Versuchsaufbauten. *Studien zur Wahrnehmung, Verarbeitung und Usability von physikbezogenen Experimenten am Bildschirm und in der Realität*
ISBN 978-3-8325-2548-4 47.50 EUR
- 105 Verena Tobias: Newton'sche Mechanik im Anfangsunterricht. *Die Wirksamkeit einer Einführung über die zweidimensionale Dynamik auf das Lehren und Lernen*
ISBN 978-3-8325-2558-3 54.00 EUR
- 106 Christian Rogge: Entwicklung physikalischer Konzepte in aufgabenbasierten Lernumgebungen
ISBN 978-3-8325-2574-3 45.00 EUR
- 107 Mathias Ropohl: Modellierung von Schülerkompetenzen im Basiskonzept Chemische Reaktion. *Entwicklung und Analyse von Testaufgaben*
ISBN 978-3-8325-2609-2 36.50 EUR
- 108 Christoph Kulgemeyer: Physikalische Kommunikationskompetenz. *Modellierung und Diagnostik*
ISBN 978-3-8325-2674-0 44.50 EUR
- 109 Jennifer Olszewski: The Impact of Physics Teachers' Pedagogical Content Knowledge on Teacher Actions and Student Outcomes
ISBN 978-3-8325-2680-1 33.50 EUR
- 110 Annika Ohle: Primary School Teachers' Content Knowledge in Physics and its Impact on Teaching and Students' Achievement
ISBN 978-3-8325-2684-9 36.50 EUR
- 111 Susanne Mannel: Assessing scientific inquiry. *Development and evaluation of a test for the low-performing stage*
ISBN 978-3-8325-2761-7 40.00 EUR
- 112 Michael Plomer: Physik physiologisch passend praktiziert. *Eine Studie zur Lernwirksamkeit von traditionellen und adressatenspezifischen Physikpraktika für die Physiologie*
ISBN 978-3-8325-2804-1 34.50 EUR
- 113 Alexandra Schulz: Experimentierspezifische Qualitätsmerkmale im Chemieunterricht. *Eine Videostudie*
ISBN 978-3-8325-2817-1 40.00 EUR
- 114 Franz Boczianowski: Eine empirische Untersuchung zu Vektoren im Physikunterricht der Mittelstufe
ISBN 978-3-8325-2843-0 39.50 EUR
- 115 Maria Ploog: Internetbasiertes Lernen durch Textproduktion im Fach Physik
ISBN 978-3-8325-2853-9 39.50 EUR

- 116 Anja Dhein: Lernen in Explorier- und Experimentiersituationen. *Eine explorative Studie zu Bedeutungsentwicklungsprozessen bei Kindern im Alter zwischen 4 und 6 Jahren*
ISBN 978-3-8325-2859-1 45.50 EUR
- 117 Irene Neumann: Beyond Physics Content Knowledge. *Modeling Competence Regarding Nature of Scientific Inquiry and Nature of Scientific Knowledge*
ISBN 978-3-8325-2880-5 37.00 EUR
- 118 Markus Emden: Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens. *Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I*
ISBN 978-3-8325-2867-6 38.00 EUR
- 119 Birgit Hofmann: Analyse von Blickbewegungen von Schülern beim Lesen von physikbezogenen Texten mit Bildern. *Eye Tracking als Methodenwerkzeug in der physikdidaktischen Forschung*
ISBN 978-3-8325-2925-3 59.00 EUR
- 120 Rebecca Knobloch: Analyse der fachinhaltlichen Qualität von Schüleräußerungen und deren Einfluss auf den Lernerfolg. *Eine Videostudie zu kooperativer Kleingruppenarbeit*
ISBN 978-3-8325-3006-8 36.50 EUR
- 121 Julia Hostenbach: Entwicklung und Prüfung eines Modells zur Beschreibung der Bewertungskompetenz im Chemieunterricht
ISBN 978-3-8325-3013-6 38.00 EUR
- 122 Anna Windt: Naturwissenschaftliches Experimentieren im Elementarbereich. *Evaluation verschiedener Lernsituationen*
ISBN 978-3-8325-3020-4 43.50 EUR
- 123 Eva Kölbach: Kontexteinflüsse beim Lernen mit Lösungsbeispielen
ISBN 978-3-8325-3025-9 38.50 EUR
- 124 Anna Lau: Passung und vertikale Vernetzung im Chemie- und Physikunterricht
ISBN 978-3-8325-3021-1 36.00 EUR
- 125 Jan Lamprecht: Ausbildungswege und Komponenten professioneller Handlungskompetenz. *Vergleich von Quereinsteigern mit Lehramtsabsolventen für Gymnasien im Fach Physik*
ISBN 978-3-8325-3035-8 38.50 EUR
- 126 Ulrike Böhm: Förderung von Verstehensprozessen unter Einsatz von Modellen
ISBN 978-3-8325-3042-6 41.00 EUR
- 127 Sabrina Dollny: Entwicklung und Evaluation eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Chemielehrkräften
ISBN 978-3-8325-3046-4 37.00 EUR
- 128 Monika Zimmermann: Naturwissenschaftliche Bildung im Kindergarten. *Eine integrative Längsschnittstudie zur Kompetenzentwicklung von Erzieherinnen*
ISBN 978-3-8325-3053-2 54.00 EUR

- 129 Ulf Saballus: Über das Schlussfolgern von Schülerinnen und Schülern zu öffentlichen Kontroversen mit naturwissenschaftlichem Hintergrund. *Eine Fallstudie*
ISBN 978-3-8325-3086-0 39.50 EUR
- 130 Olaf Krey: Zur Rolle der Mathematik in der Physik. *Wissenschaftstheoretische Aspekte und Vorstellungen Physiklernender*
ISBN 978-3-8325-3101-0 46.00 EUR
- 131 Angelika Wolf: Zusammenhänge zwischen der Eigenständigkeit im Physikunterricht, der Motivation, den Grundbedürfnissen und dem Lernerfolg von Schülern
ISBN 978-3-8325-3161-4 45.00 EUR
- 132 Johannes Börlin: Das Experiment als Lerngelegenheit. *Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität*
ISBN 978-3-8325-3170-6 45.00 EUR
- 133 Olaf Uhden: Mathematisches Denken im Physikunterricht. *Theorieentwicklung und Problemanalyse*
ISBN 978-3-8325-3170-6 45.00 EUR
- 134 Christoph Gut: Modellierung und Messung experimenteller Kompetenz. *Analyse eines large-scale Experimentiertests*
ISBN 978-3-8325-3213-0 40.00 EUR
- 135 Antonio Rueda: Lernen mit ExploMultimedial in kolumbianischen Schulen. *Analyse von kurzzeitigen Lernprozessen und der Motivation beim länderübergreifenden Einsatz einer deutschen computergestützten multimedialen Lernumgebung für den naturwissenschaftlichen Unterricht*
ISBN 978-3-8325-3218-5 45.50 EUR
- 136 Krisztina Berger: Bilder, Animationen und Notizen. *Empirische Untersuchung zur Wirkung einfacher visueller Repräsentationen und Notizen auf den Wissenserwerb in der Optik*
ISBN 978-3-8325-3238-3 41.50 EUR
- 137 Antony Crossley: Untersuchung des Einflusses unterschiedlicher physikalischer Konzepte auf den Wissenserwerb in der Thermodynamik der Sekundarstufe I
ISBN 978-3-8325-3275-8 40.00 EUR
- 138 Tobias Viering: Entwicklung physikalischer Kompetenz in der Sekundarstufe I. *Validierung eines Kompetenzentwicklungsmodells für das Energiekonzept im Bereich Fachwissen*
ISBN 978-3-8325-3277-2 37.00 EUR
- 139 Nico Schreiber: Diagnostik experimenteller Kompetenz. *Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells*
ISBN 978-3-8325-3284-0 39.00 EUR
- 140 Sarah Hundertmark: Einblicke in kollaborative Lernprozesse. *Eine Fallstudie zur reflektierenden Zusammenarbeit unterstützt durch die Methoden Concept Mapping und Lernbegleitbogen*
ISBN 978-3-8325-3251-2 43.00 EUR

- 141 Ronny Scherer: Analyse der Struktur, Messinvarianz und Ausprägung komplexer Problemlösekompetenz im Fach Chemie. *Eine Querschnittstudie in der Sekundarstufe I und am Übergang zur Sekundarstufe II*
ISBN 978-3-8325-3312-0 43.00 EUR
- 142 Patricia Heitmann: Bewertungskompetenz im Rahmen naturwissenschaftlicher Problemlöseprozesse. *Modellierung und Diagnose der Kompetenzen Bewertung und analytisches Problemlösen für das Fach Chemie*
ISBN 978-3-8325-3314-4 37.00 EUR
- 143 Jan Fleischhauer: Wissenschaftliches Argumentieren und Entwicklung von Konzepten beim Lernen von Physik
ISBN 978-3-8325-3325-0 35.00 EUR
- 144 Nermin Özcan: Zum Einfluss der Fachsprache auf die Leistung im Fach Chemie. *Eine Förderstudie zur Fachsprache im Chemieunterricht*
ISBN 978-3-8325-3328-1 36.50 EUR
- 145 Helena van Vorst: Kontextmerkmale und ihr Einfluss auf das Schülerinteresse im Fach Chemie
ISBN 978-3-8325-3321-2 38.50 EUR
- 146 Janine Cappell: Fachspezifische Diagnosekompetenz angehender Physiklehrkräfte in der ersten Ausbildungsphase
ISBN 978-3-8325-3356-4 38.50 EUR
- 147 Susanne Bley: Förderung von Transferprozessen im Chemieunterricht
ISBN 978-3-8325-3407-3 40.50 EUR
- 148 Cathrin Blaes: Die übungsgestützte Lehrerrepräsentation im Chemieunterricht der Sekundarstufe I. *Evaluation der Effektivität*
ISBN 978-3-8325-3409-7 43.50 EUR
- 149 Julia Suckut: Die Wirksamkeit von piko-OWL als Lehrerfortbildung. Eine Evaluation zum Projekt *Physik im Kontext* in Fallstudien
ISBN 978-3-8325-3440-0 45.00 EUR
- 150 Alexandra Dorschu: Die Wirkung von Kontexten in Physikkompetenztestaufgaben
ISBN 978-3-8325-3446-2 37.00 EUR
- 151 Jochen Scheid: Multiple Repräsentationen, Verständnis physikalischer Experimente und kognitive Aktivierung: *Ein Beitrag zur Entwicklung der Aufgabenkultur*
ISBN 978-3-8325-3449-3 49.00 EUR
- 152 Tim Plasa: Die Wahrnehmung von Schülerlaboren und Schülerforschungszentren
ISBN 978-3-8325-3483-7 35.50 EUR
- 153 Felix Schoppmeier: Physikkompetenz in der gymnasialen Oberstufe. *Entwicklung und Validierung eines Kompetenzstrukturmodells für den Kompetenzbereich Umgang mit Fachwissen*
ISBN 978-3-8325-3502-5 36.00 EUR

- 154 Katharina Groß: Experimente alternativ dokumentieren. *Eine qualitative Studie zur Förderung der Diagnose- und Differenzierungskompetenz in der Chemielehrerbildung*
ISBN 978-3-8325-3508-7 43.50 EUR
- 155 Barbara Hank: Konzeptwandelprozesse im Anfangsunterricht Chemie. *Eine quasixperimentelle Längsschnittstudie*
ISBN 978-3-8325-3519-3 38.50 EUR
- 156 Katja Freyer: Zum Einfluss von Studieneingangsvoraussetzungen auf den Studienerfolg Erstsemesterstudierender im Fach Chemie
ISBN 978-3-8325-3544-5 38.00 EUR
- 157 Alexander Rachel: Auswirkungen instruktorischer Hilfen bei der Einführung des (Ferro-)Magnetismus. *Eine Vergleichsstudie in der Primar- und Sekundarstufe*
ISBN 978-3-8325-3548-3 43.50 EUR
- 158 Sebastian Ritter: Einfluss des Lerninhalts Nanogrößeneffekte auf Teilchen- und Teilchenmodellvorstellungen von Schülerinnen und Schülern
ISBN 978-3-8325-3558-2 36.00 EUR
- 159 Andrea Harbach: Problemorientierung und Vernetzung in kontextbasierten Lernaufgaben
ISBN 978-3-8325-3564-3 39.00 EUR
- 160 David Obst: Interaktive Tafeln im Physikunterricht. *Entwicklung und Evaluation einer Lehrerfortbildung*
ISBN 978-3-8325-3582-7 40.50 EUR
- 161 Sophie Kirschner: Modellierung und Analyse des Professionswissens von Physiklehrkräften
ISBN 978-3-8325-3601-5 35.00 EUR
- 162 Katja Stief: Selbstregulationsprozesse und Hausaufgabenmotivation im Chemieunterricht
ISBN 978-3-8325-3631-2 34.00 EUR
- 163 Nicola Meschede: Professionelle Wahrnehmung der inhaltlichen Strukturierung im naturwissenschaftlichen Grundschulunterricht. *Theoretische Beschreibung und empirische Erfassung*
ISBN 978-3-8325-3668-8 37.00 EUR
- 164 Johannes Maximilian Barth: Experimentieren im Physikunterricht der gymnasialen Oberstufe. *Eine Rekonstruktion übergeordneter Einbettungsstrategien*
ISBN 978-3-8325-3681-7 39.00 EUR
- 165 Sandra Lein: Das Betriebspraktikum in der Lehrerbildung. *Eine Untersuchung zur Förderung der Wissenschafts- und Technikbildung im allgemeinbildenden Unterricht*
ISBN 978-3-8325-3698-5 40.00 EUR
- 166 Veranika Maiseyenko: Modellbasiertes Experimentieren im Unterricht. *Praxistauglichkeit und Lernwirkungen*
ISBN 978-3-8325-3708-1 38.00 EUR

- 167 Christoph Stolzenberger: Der Einfluss der didaktischen Lernumgebung auf das Erreichen geforderter Bildungsziele am Beispiel der W- und P-Seminare im Fach Physik
ISBN 978-3-8325-3708-1 38.00 EUR
- 168 Pia Altenburger: Mehrebenenregressionsanalysen zum Physiklernen im Sachunterricht der Primarstufe. *Ergebnisse einer Evaluationsstudie.*
ISBN 978-3-8325-3717-3 37.50 EUR
- 169 Nora Ferber: Entwicklung und Validierung eines Testinstruments zur Erfassung von Kompetenzentwicklung im Fach Chemie in der Sekundarstufe I
ISBN 978-3-8325-3727-2 39.50 EUR
- 170 Anita Stender: Unterrichtsplanung: Vom Wissen zum Handeln.
Theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung
ISBN 978-3-8325-3750-0 41.50 EUR
- 171 Jenna Koenen: Entwicklung und Evaluation von experimentunterstützten Lösungsbeispielen zur Förderung naturwissenschaftlich-experimenteller Arbeitsweisen
ISBN 978-3-8325-3785-2 43.00 EUR
- 172 Teresa Henning: Empirische Untersuchung kontextorientierter Lernumgebungen in der Hochschuldidaktik. *Entwicklung und Evaluation kontextorientierter Aufgaben in der Studieneingangsphase für Fach- und Nebenfachstudierende der Physik*
ISBN 978-3-8325-3801-9 43.00 EUR
- 173 Alexander Pusch: Fachspezifische Instrumente zur Diagnose und individuellen Förderung von Lehramtsstudierenden der Physik
ISBN 978-3-8325-3829-3 38.00 EUR
- 174 Christoph Vogelsang: Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften. *Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz*
ISBN 978-3-8325-3846-0 50.50 EUR
- 175 Ingo Brebeck: Selbstreguliertes Lernen in der Studieneingangsphase im Fach Chemie
ISBN 978-3-8325-3859-0 37.00 EUR
- 176 Axel Eghtessad: Merkmale und Strukturen von Professionalisierungsprozessen in der ersten und zweiten Phase der Chemielehrerbildung. *Eine empirisch-qualitative Studie mit niedersächsischen Fachleiter_innen der Sekundarstufenlehrämter*
ISBN 978-3-8325-3861-3 45.00 EUR
- 177 Andreas Nehring: Wissenschaftliche Denk- und Arbeitsweisen im Fach Chemie. Eine kompetenzorientierte Modell- und Testentwicklung für den Bereich der Erkenntnisgewinnung
ISBN 978-3-8325-3872-9 39.50 EUR
- 178 Maike Schmidt: Professionswissen von Sachunterrichtslehrkräften. Zusammenhangsanalyse zur Wirkung von Ausbildungshintergrund und Unterrichtserfahrung auf das fachspezifische Professionswissen im Unterrichtsinhalt „Verbrennung“
ISBN 978-3-8325-3907-8 38.50 EUR

- 179 Jan Winkelmann: Auswirkungen auf den Fachwissenszuwachs und auf affektive Schülermerkmale durch Schüler- und Demonstrationsexperimente im Physikunterricht
ISBN 978-3-8325-3915-3 41.00 EUR
- 180 Iwen Kobow: Entwicklung und Validierung eines Testinstrumentes zur Erfassung der Kommunikationskompetenz im Fach Chemie
ISBN 978-3-8325-3927-6 34.50 EUR
- 181 Yvonne Gramzow: Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion
ISBN 978-3-8325-3931-3 42.50 EUR
- 182 Evelin Schröter: Entwicklung der Kompetenzerwartung durch Lösen physikalischer Aufgaben einer multimedialen Lernumgebung
ISBN 978-3-8325-3975-7 54.50 EUR
- 183 Inga Kallweit: Effektivität des Einsatzes von Selbsteinschätzungsbögen im Chemieunterricht der Sekundarstufe I. *Individuelle Förderung durch selbstreguliertes Lernen*
ISBN 978-3-8325-3965-8 44.00 EUR
- 184 Andrea Schumacher: Paving the way towards authentic chemistry teaching. *A contribution to teachers' professional development*
ISBN 978-3-8325-3976-4 48.50 EUR
- 185 David Woitkowski: Fachliches Wissen Physik in der Hochschulausbildung. *Konzeptualisierung, Messung, Niveaubildung*
ISBN 978-3-8325-3988-7 53.00 EUR
- 186 Marianne Korner: Cross-Age Peer Tutoring in Physik. *Evaluation einer Unterrichtsmethode*
ISBN 978-3-8325-3979-5 38.50 EUR
- 187 Simone Nakoinz: Untersuchung zur Verknüpfung submikroskopischer und makroskopischer Konzepte im Fach Chemie
ISBN 978-3-8325-4057-9 38.50 EUR
- 188 Sandra Anus: Evaluation individueller Förderung im Chemieunterricht. *Adaptivität von Lerninhalten an das Vorwissen von Lernenden am Beispiel des Basiskonzeptes Chemische Reaktion*
ISBN 978-3-8325-4059-3 43.50 EUR
- 189 Thomas Roßbegalle: Fachdidaktische Entwicklungsforschung zum besseren Verständnis atmosphärischer Phänomene. *Treibhauseffekt, saurer Regen und stratosphärischer Ozonabbau als Kontexte zur Vermittlung von Basiskonzepten der Chemie*
ISBN 978-3-8325-4059-3 45.50 EUR
- 190 Kathrin Steckenmesser-Sander: Gemeinsamkeiten und Unterschiede physikbezogener Handlungs-, Denk- und Lernprozesse von Mädchen und Jungen
ISBN 978-3-8325-4066-1 38.50 EUR
- 191 Cornelia Geller: Lernprozessorientierte Sequenzierung des Physikunterrichts im Zusammenhang mit Fachwissenserwerb. *Eine Videostudie in Finnland, Deutschland und der Schweiz*
ISBN 978-3-8325-4082-1 35.50 EUR

- 192 Jan Hofmann: Untersuchung des Kompetenzaufbaus von Physiklehrkräften während einer Fortbildungsmaßnahme
ISBN 978-3-8325-4104-0 38.50 EUR
- 193 Andreas Dickhäuser: Chemiespezifischer Humor. *Theoriebildung, Materialentwicklung, Evaluation*
ISBN 978-3-8325-4108-8 37.00 EUR
- 194 Stefan Korte: Die Grenzen der Naturwissenschaft als Thema des Physikunterrichts
ISBN 978-3-8325-4112-5 57.50 EUR
- 195 Carolin Hülsmann: Kurswahlmotive im Fach Chemie. Eine Studie zum Wahlverhalten und Erfolg von Schülerinnen und Schülern in der gymnasialen Oberstufe
ISBN 978-3-8325-4144-6 49.00 EUR
- 196 Caroline Körbs: Mindeststandards im Fach Chemie am Ende der Pflichtschulzeit
ISBN 978-3-8325-4148-4 34.00 EUR
- 197 Andreas Vorholzer: Wie lassen sich Kompetenzen des experimentellen Denkens und Arbeitens fördern? *Eine empirische Untersuchung der Wirkung eines expliziten und eines impliziten Instruktionsansatzes*
ISBN 978-3-8325-4194-1 37.50 EUR
- 198 Anna Katharina Schmitt: Entwicklung und Evaluation einer Chemielehrerfortbildung zum Kompetenzbereich Erkenntnisgewinnung
ISBN 978-3-8325-4228-3 39.50 EUR
- 199 Christian Maurer: Strukturierung von Lehr-Lern-Sequenzen
ISBN 978-3-8325-4247-4 36.50 EUR
- 200 Helmut Fischler, Elke Sumfleth (Hrsg.): Professionelle Kompetenz von Lehrkräften der Chemie und Physik
ISBN 978-3-8325-4523-9 34.00 EUR
- 201 Simon Zander: Lehrerfortbildung zu Basismodellen und Zusammenhänge zum Fachwissen
ISBN 978-3-8325-4248-1 35.00 EUR
- 202 Kerstin Arndt: Experimentierkompetenz erfassen.
Analyse von Prozessen und Mustern am Beispiel von Lehramtsstudierenden der Chemie
ISBN 978-3-8325-4266-5 45.00 EUR
- 203 Christian Lang: Kompetenzorientierung im Rahmen experimentalchemischer Praktika
ISBN 978-3-8325-4268-9 42.50 EUR
- 204 Eva Cauet: Testen wir relevantes Wissen? *Zusammenhang zwischen dem Professionswissen von Physiklehrkräften und gutem und erfolgreichem Unterrichten*
ISBN 978-3-8325-4276-4 39.50 EUR
- 205 Patrick Löffler: Modellanwendung in Problemlöseaufgaben. *Wie wirkt Kontext?*
ISBN 978-3-8325-4303-7 35.00 EUR

- 206 Carina Gehlen: Kompetenzstruktur naturwissenschaftlicher Erkenntnisgewinnung im Fach Chemie
ISBN 978-3-8325-4318-1 43.00 EUR
- 207 Lars Oettinghaus: Lehrerüberzeugungen und physikbezogenes Professionswissen. *Vergleich von Absolventinnen und Absolventen verschiedener Ausbildungswege im Physikreferendariat*
ISBN 978-3-8325-4319-8 38.50 EUR
- 208 Jennifer Petersen: Zum Einfluss des Merkmals Humor auf die Gesundheitsförderung im Chemieunterricht der Sekundarstufe I. *Eine Interventionsstudie zum Thema Sonnenschutz*
ISBN 978-3-8325-4348-8 40.00 EUR
- 209 Philipp Straube: Modellierung und Erfassung von Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung bei (Lehramts-) Studierenden im Fach Physik
ISBN 978-3-8325-4351-8 35.50 EUR
- 210 Martin Dickmann: Messung von Experimentierfähigkeiten. *Validierungsstudien zur Qualität eines computerbasierten Testverfahrens*
ISBN 978-3-8325-4356-3 41.00 EUR
- 211 Markus Bohlmann: Science Education. Empirie, Kulturen und Mechanismen der Didaktik der Naturwissenschaften
ISBN 978-3-8325-4377-8 44.00 EUR
- 212 Martin Draude: Die Kompetenz von Physiklehrkräften, Schwierigkeiten von Schülerinnen und Schülern beim eigenständigen Experimentieren zu diagnostizieren
ISBN 978-3-8325-4382-2 37.50 EUR
- 213 Henning Rode: Prototypen evidenzbasierten Physikunterrichts. *Zwei empirische Studien zum Einsatz von Feedback und Blackboxes in der Sekundarstufe*
ISBN 978-3-8325-4389-1 42.00 EUR
- 214 Jan-Henrik Kechel: Schülerschwierigkeiten beim eigenständigen Experimentieren. *Eine qualitative Studie am Beispiel einer Experimentieraufgabe zum Hooke'schen Gesetz*
ISBN 978-3-8325-4392-1 55.00 EUR
- 215 Katharina Fricke: Classroom Management and its Impact on Lesson Outcomes in Physics. *A multi-perspective comparison of teaching practices in primary and secondary schools*
ISBN 978-3-8325-4394-5 40.00 EUR
- 216 Hannes Sander: Orientierungen von Jugendlichen beim Urteilen und Entscheiden in Kontexten nachhaltiger Entwicklung. *Eine rekonstruktive Perspektive auf Bewertungskompetenz in der Didaktik der Naturwissenschaft*
ISBN 978-3-8325-4434-8 46.00 EUR
- 217 Inka Haak: Maßnahmen zur Unterstützung kognitiver und metakognitiver Prozesse in der Studieneingangsphase. *Eine Design-Based-Research-Studie zum universitären Lernzentrum Physiktreff*
ISBN 978-3-8325-4437-9 46.50 EUR

- 218 Martina Brandenburger: Was beeinflusst den Erfolg beim Problemlösen in der Physik?
Eine Untersuchung mit Studierenden
ISBN 978-3-8325-4409-6 42.50 EUR
- 219 Corinna Helms: Entwicklung und Evaluation eines Trainings zur Verbesserung der Erklärqualität von Schülerinnen und Schülern im Gruppenpuzzle
ISBN 978-3-8325-4454-6 42.50 EUR
- 220 Viktoria Rath: Diagnostische Kompetenz von angehenden Physiklehrkräften. *Modellierung, Testinstrumentenentwicklung und Erhebung der Performanz bei der Diagnose von Schülervorstellungen in der Mechanik*
ISBN 978-3-8325-4456-0 42.50 EUR
- 221 Janne Krüger: Schülerperspektiven auf die zeitliche Entwicklung der Naturwissenschaften
ISBN 978-3-8325-4457-7 45.50 EUR
- 222 Stefan Mutke: Das Professionswissen von Chemiereferendarinnen und -referendaren in Nordrhein-Westfalen. *Eine Längsschnittstudie*
ISBN 978-3-8325-4458-4 37.50 EUR
- 223 Sebastian Habig: Systematisch variierte Kontextaufgaben und ihr Einfluss auf kognitive und affektive Schülerfaktoren
ISBN 978-3-8325-4467-6 40.50 EUR
- 224 Sven Liepertz: Zusammenhang zwischen dem Professionswissen von Physiklehrkräften, dem sachstrukturellen Angebot des Unterrichts und der Schülerleistung
ISBN 978-3-8325-4480-5 34.00 EUR
- 225 Elina Platova: Optimierung eines Laborpraktikums durch kognitive Aktivierung
ISBN 978-3-8325-4481-2 39.00 EUR
- 226 Tim Reschke: Lesegeschichten im Chemieunterricht der Sekundarstufe I zur Unterstützung von situationalem Interesse und Lernerfolg
ISBN 978-3-8325-4487-4 41.00 EUR
- 227 Lena Mareike Walper: Entwicklung der physikbezogenen Interessen und selbstbezogenen Kognitionen von Schülerinnen und Schülern in der Übergangsphase von der Primar- in die Sekundarstufe. *Eine Längsschnittanalyse vom vierten bis zum siebten Schuljahr*
ISBN 978-3-8325-4495-9 43.00 EUR
- 228 Stefan Anthofer: Förderung des fachspezifischen Professionswissens von Chemielehramtsstudierenden
ISBN 978-3-8325-4498-0 39.50 EUR
- 229 Marcel Bullinger: Handlungsorientiertes Physiklernen mit instruierten Selbsterklärungen in der Primarstufe. *Eine experimentelle Laborstudie*
ISBN 978-3-8325-4504-8 44.00 EUR
- 230 Thomas Amenda: Bedeutung fachlicher Elementarisierungen für das Verständnis der Kinematik
ISBN 978-3-8325-4531-4 43.50 EUR

- 231 Sabrina Milke: Beeinflusst *Priming* das Physiklernen?
Eine empirische Studie zum Dritten Newtonschen Axiom
ISBN 978-3-8325-4549-4 42.00 EUR
- 232 Corinna Erfmann: Ein anschaulicher Weg zum Verständnis der elektromagnetischen Induktion. *Evaluation eines Unterrichtsvorschlags und Validierung eines Leistungsdiagnoseinstruments*
ISBN 978-3-8325-4550-5 49.50 EUR
- 233 Hanne Rautenstrauch: Erhebung des (Fach-)Sprachstandes bei Lehramtsstudierenden im Kontext des Faches Chemie
ISBN 978-3-8325-4556-7 40.50 EUR
- 234 Tobias Klug: Wirkung kontextorientierter physikalischer Praktikumsversuche auf Lernprozesse von Studierenden der Medizin
ISBN 978-3-8325-4558-1 37.00 EUR
- 235 Mareike Bohrmann: Zur Förderung des Verständnisses der Variablenkontrolle im naturwissenschaftlichen Sachunterricht
ISBN 978-3-8325-4559-8 52.00 EUR
- 236 Anja Schödl: FALKO-Physik – Fachspezifische Lehrerkompetenzen im Fach Physik. *Entwicklung und Validierung eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Physiklehrkräften*
ISBN 978-3-8325-4553-6 40.50 EUR
- 237 Hilda Scheuermann: Entwicklung und Evaluation von Unterstützungsmaßnahmen zur Förderung der Variablenkontrollstrategie beim Planen von Experimenten
ISBN 978-3-8325-4568-0 39.00 EUR
- 238 Christian G. Strippel: Naturwissenschaftliche Erkenntnisgewinnung an chemischen Inhalten vermitteln. *Konzeption und empirische Untersuchung einer Ausstellung mit Experimentierstation*
ISBN 978-3-8325-4577-2 41.50 EUR
- 239 Sarah Rau: Durchführung von Sachunterricht im Vorbereitungsdienst. *Eine längsschnittliche, videobasierte Unterrichtsanalyse*
ISBN 978-3-8325-4579-6 46.00 EUR
- 240 Thomas Plotz: Lernprozesse zu nicht-sichtbarer Strahlung. *Empirische Untersuchungen in der Sekundarstufe 2*
ISBN 978-3-8325-4624-3 39.50 EUR
- 241 Wolfgang Aschauer: Elektrische und magnetische Felder. *Eine empirische Studie zu Lernprozessen in der Sekundarstufe II*
ISBN 978-3-8325-4625-0 50.00 EUR
- 242 Anna Donhauser: Didaktisch rekonstruierte Materialwissenschaft. *Aufbau und Konzeption eines Schülerlabors für den Exzellenzcluster Engineering of Advanced Materials*
ISBN 978-3-8325-4636-6 39.00 EUR

- 243 Katrin Schüßler: Lernen mit Lösungsbeispielen im Chemieunterricht. *Einflüsse auf Lernerfolg, kognitive Belastung und Motivation*
ISBN 978-3-8325-4640-3 42.50 EUR
- 244 Timo Fleischer: Untersuchung der chemischen Fachsprache unter besonderer Berücksichtigung chemischer Repräsentationen
ISBN 978-3-8325-4642-7 46.50 EUR
- 245 Rosina Steininger: Concept Cartoons als Stimuli für Kleingruppendiskussionen im Chemieunterricht. *Beschreibung und Analyse einer komplexen Lerngelegenheit*
ISBN 978-3-8325-4647-2 39.00 EUR
- 246 Daniel Rehfeldt: Erfassung der Lehrqualität naturwissenschaftlicher Experimentalpraktika
ISBN 978-3-8325-4590-1 40.00 EUR
- 247 Sandra Puddu: Implementing Inquiry-based Learning in a Diverse Classroom: Investigating Strategies of Scaffolding and Students' Views of Scientific Inquiry
ISBN 978-3-8325-4591-8 35.50 EUR
- 248 Markus Bliersbach: Kreativität in der Chemie. *Erhebung und Förderung der Vorstellungen von Chemielehramtsstudierenden*
ISBN 978-3-8325-4593-2 44.00 EUR
- 249 Lennart Kimpel: Aufgaben in der Allgemeinen Chemie. *Zum Zusammenspiel von chemischem Verständnis und Rechenfähigkeit*
ISBN 978-3-8325-4618-2 36.00 EUR
- 250 Louise Bindel: Effects of integrated learning: explicating a mathematical concept in inquiry-based science camps
ISBN 978-3-8325-4655-7 37.50 EUR
- 251 Michael Wenzel: Computereinsatz in Schule und Schülerlabor. *Einstellung von Physiklehrkräften zu Neuen Medien*
ISBN 978-3-8325-4659-5 38.50 EUR
- 252 Laura Muth: Einfluss der Auswertephase von Experimenten im Physikunterricht. *Ergebnisse einer Interventionsstudie zum Zuwachs von Fachwissen und experimenteller Kompetenz von Schülerinnen und Schülern*
ISBN 978-3-8325-4675-5 36.50 EUR
- 253 Annika Fricke: Interaktive Skripte im Physikalischen Praktikum. *Entwicklung und Evaluation von Hypermedien für die Nebenfachausbildung*
ISBN 978-3-8325-4676-2 41.00 EUR
- 254 Julia Haase: Selbstbestimmtes Lernen im naturwissenschaftlichen Sachunterricht. *Eine empirische Interventionsstudie mit Fokus auf Feedback und Kompetenzerleben*
ISBN 978-3-8325-4685-4 38.50 EUR
- 255 Antje J. Heine: Was ist Theoretische Physik? *Eine wissenschaftstheoretische Betrachtung und Rekonstruktion von Vorstellungen von Studierenden und Dozenten über das Wesen der Theoretischen Physik*
ISBN 978-3-8325-4691-5 46.50 EUR

- 256 Claudia Meinhardt: Entwicklung und Validierung eines Testinstruments zu Selbstwirksamkeitserwartungen von (angehenden) Physiklehrkräften in physikdidaktischen Handlungsfeldern
ISBN 978-3-8325-4712-7 47.00 EUR
- 257 Ann-Kathrin Schlüter: Professionalisierung angehender Chemielehrkräfte für einen Gemeinsamen Unterricht
ISBN 978-3-8325-4713-4 53.50 EUR
- 258 Stefan Richtberg: Elektronenbahnen in Feldern. Konzeption und Evaluation einer webbasierten Lernumgebung
ISBN 978-3-8325-4723-3 49.00 EUR
- 259 Jan-Philipp Burde: Konzeption und Evaluation eines Unterrichtskonzepts zu einfachen Stromkreisen auf Basis des Elektronengasmodells
ISBN 978-3-8325-4726-4 57.50 EUR
- 260 Frank Finkenberg: Flipped Classroom im Physikunterricht
ISBN 978-3-8325-4737-4 42.50 EUR
- 261 Florian Treisch: Die Entwicklung der Professionellen Unterrichtswahrnehmung im Lehr-Lern-Labor Seminar
ISBN 978-3-8325-4741-4 41.50 EUR
- 262 Desiree Mayr: Strukturiertheit des experimentellen naturwissenschaftlichen Problemlöseprozesses
ISBN 978-3-8325-4757-8 37.00 EUR
- 263 Katrin Weber: Entwicklung und Validierung einer Learning Progression für das Konzept der chemischen Reaktion in der Sekundarstufe I
ISBN 978-3-8325-4762-2 48.50 EUR
- 264 Hauke Bartels: Entwicklung und Bewertung eines performanznahen Videovignetten-tests zur Messung der Erklärfähigkeit von Physiklehrkräften
ISBN 978-3-8325-4804-9 37.00 EUR
- 265 Karl Marniok: Zum Wesen von Theorien und Gesetzen in der Chemie. *Begriffsanalyse und Förderung der Vorstellungen von Lehramtsstudierenden*
ISBN 978-3-8325-4805-6 42.00 EUR
- 266 Marisa Holzapfel: Fachspezifischer Humor als Methode in der Gesundheitsbildung im Übergang von der Primarstufe zur Sekundarstufe I
ISBN 978-3-8325-4808-7 50.00 EUR
- 267 Anna Stolz: Die Auswirkungen von Experimentiersituationen mit unterschiedlichem Öffnungsgrad auf Leistung und Motivation der Schülerinnen und Schüler
ISBN 978-3-8325-4781-3 38.00 EUR
- 268 Nina Ulrich: Interaktive Lernaufgaben in dem digitalen Schulbuch eChemBook. *Einfluss des Interaktivitätsgrads der Lernaufgaben und des Vorwissens der Lernenden auf den Lernerfolg*
ISBN 978-3-8325-4814-8 43.50 EUR

- 269 Kim-Alessandro Weber: Quantenoptik in der Lehrerfortbildung. *Ein bedarfsgeprägtes Fortbildungskonzept zum Quantenobjekt Photon mit Realexperimenten*
ISBN 978-3-8325-4792-9 55.00 EUR
- 270 Nina Skorsetz: Empathisierer und Systematisierer im Vorschulalter. *Eine Fragebogen- und Videostudie zur Motivation, sich mit Naturphänomenen zu beschäftigen*
ISBN 978-3-8325-4825-4 43.50 EUR
- 271 Franziska Kehne: Analyse des Transfers von kontextualisiert erworbenem Wissen im Fach Chemie
ISBN 978-3-8325-4846-9 45.00 EUR
- 272 Markus Elsholz: Das akademische Selbstkonzept angehender Physiklehrkräfte als Teil ihrer professionellen Identität. *Dimensionalität und Veränderung während einer zentralen Praxisphase*
ISBN 978-3-8325-4857-5 37.50 EUR
- 273 Joachim Müller: Studienerfolg in der Physik. *Zusammenhang zwischen Modellierungskompetenz und Studienerfolg*
ISBN 978-3-8325-4859-9 35.00 EUR
- 274 Jennifer Dörscheln: Organische Leuchtdioden. *Implementation eines innovativen Themas in den Chemieunterricht*
ISBN 978-3-8325-4865-0 59.00 EUR
- 275 Stephanie Strelow: Beliefs von Studienanfängern des Kombi-Bachelors Physik über die Natur der Naturwissenschaften
ISBN 978-3-8325-4881-0 40.50 EUR
- 276 Dennis Jaeger: Kognitive Belastung und aufgabenspezifische sowie personenspezifische Einflussfaktoren beim Lösen von Physikaufgaben
ISBN 978-3-8325-4928-2 50.50 EUR
- 277 Vanessa Fischer: Der Einfluss von Interesse und Motivation auf die Messung von Fach- und Bewertungskompetenz im Fach Chemie
ISBN 978-3-8325-4933-6 39.00 EUR
- 278 René Dohrmann: Professionsbezogene Wirkungen einer Lehr-Lern-Labor-Veranstaltung. *Eine multimethodische Studie zu den professionsbezogenen Wirkungen einer Lehr-Lern-Labor-Blockveranstaltung auf Studierende der Bachelorstudiengänge Lehramt Physik und Grundschulpädagogik (Sachunterricht)*
ISBN 978-3-8325-4958-9 40.00 EUR
- 279 Meike Bergs: Can We Make Them Use These Strategies? *Fostering Inquiry-Based Science Learning Skills with Physical and Virtual Experimentation Environments*
ISBN 978-3-8325-4962-6 39.50 EUR
- 280 Marie-Therese Hauerstein: Untersuchung zur Effektivität von Strukturierung und Binnendifferenzierung im Chemieunterricht der Sekundarstufe I. *Evaluation der Strukturierungshilfe Lernleiter*
ISBN 978-3-8325-4982-4 42.50 EUR

- 281 Verena Zucker: Erkennen und Beschreiben von formativem Assessment im naturwissenschaftlichen Grundschulunterricht. *Entwicklung eines Instruments zur Erfassung von Teilfähigkeiten der professionellen Wahrnehmung von Lehramtsstudierenden*
ISBN 978-3-8325-4991-6 38.00 EUR
- 282 Victoria Telser: Erfassung und Förderung experimenteller Kompetenz von Lehrkräften im Fach Chemie
ISBN 978-3-8325-4996-1 50.50 EUR
- 283 Kristine Tschirschky: Entwicklung und Evaluation eines gedächtnisorientierten Aufgabendesigns für Physikaufgaben
ISBN 978-3-8325-5002-8 42.50 EUR
- 284 Thomas Elert: Course Success in the Undergraduate General Chemistry Lab
ISBN 978-3-8325-5004-2 41.50 EUR
- 285 Britta Kalthoff: Explizit oder implizit? *Untersuchung der Lernwirksamkeit verschiedener fachmethodischer Instruktionen im Hinblick auf fachmethodische und fachinhaltliche Fähigkeiten von Sachunterrichtsstudierenden*
ISBN 978-3-8325-5013-4 37.50 EUR
- 286 Thomas Dickmann: Visuelles Modellverständnis und Studienerfolg in der Chemie. *Zwei Seiten einer Medaille*
ISBN 978-3-8325-5016-5 44.00 EUR
- 287 Markus Sebastian Feser: Physiklehrkräfte korrigieren Schülertexte. *Eine Explorationsstudie zur fachlich-konzeptuellen und sprachlichen Leistungsfeststellung und -beurteilung im Physikunterricht*
ISBN 978-3-8325-5020-2 49.00 EUR
- 288 Matylda Dudzinska: Lernen mit Beispielaufgaben und Feedback im Physikunterricht der Sekundarstufe 1. *Energieerhaltung zur Lösung von Aufgaben nutzen*
ISBN 978-3-8325-5025-7 47.00 EUR
- 289 Ines Sonnenschein: Naturwissenschaftliche Denk- und Arbeitsprozesse Studierender im Labor
ISBN 978-3-8325-5033-2 52.00 EUR
- 290 Florian Simon: Der Einfluss von Betreuung und Betreuenden auf die Wirksamkeit von Schülerlaborbesuchen. *Eine Zusammenhangsanalyse von Betreuungsqualität, Betreuermerkmalen und Schülerlaborzielen sowie Replikationsstudie zur Wirksamkeit von Schülerlaborbesuchen*
ISBN 978-3-8325-5036-3 49.50 EUR
- 291 Marie-Annette Geyer: Physikalisch-mathematische Darstellungswechsel funktionaler Zusammenhänge. *Das Vorgehen von SchülerInnen der Sekundarstufe 1 und ihre Schwierigkeiten*
ISBN 978-3-8325-5047-9 46.50 EUR
- 292 Susanne Digel: Messung von Modellierungskompetenz in Physik. *Theoretische Herleitung und empirische Prüfung eines Kompetenzmodells physikspezifischer Modellierungskompetenz*
ISBN 978-3-8325-5055-4 41.00 EUR

- 293 Sönke Janssen: Angebots-Nutzungs-Prozesse eines Schülerlabors analysieren und gestalten. *Ein design-based research Projekt*
ISBN 978-3-8325-5065-3 57.50 EUR
- 294 Knut Wille: Der Productive Failure Ansatz als Beitrag zur Weiterentwicklung der Aufgabenkultur
ISBN 978-3-8325-5074-5 49.00 EUR
- 295 Lisanne Kraeva: Problemlösestrategien von Schülerinnen und Schülern diagnostizieren
ISBN 978-3-8325-5110-0 59.50 EUR
- 296 Jenny Lorentzen: Entwicklung und Evaluation eines Lernangebots im Lehramtsstudium Chemie zur Förderung von Vernetzungen innerhalb des fachbezogenen Professionswissens
ISBN 978-3-8325-5120-9 39.50 EUR
- 297 Micha Winkelmann: Lernprozesse in einem Schülerlabor unter Berücksichtigung individueller naturwissenschaftlicher Interessenstrukturen
ISBN 978-3-8325-5147-6 48.50 EUR
- 298 Carina Wöhlke: Entwicklung und Validierung eines Instruments zur Erfassung der professionellen Unterrichtswahrnehmung angehender Physiklehrkräfte
ISBN 978-3-8325-5149-0 43.00 EUR
- 299 Thomas Schubatzky: Das Amalgam Anfangs-Elektrizitätslehreunterricht. *Eine multiperspektivische Betrachtung in Deutschland und Österreich*
ISBN 978-3-8325-5159-9 50.50 EUR
- 300 Amany Annaggar: A Design Framework for Video Game-Based Gamification Elements to Assess Problem-solving Competence in Chemistry Education
ISBN 978-3-8325-5150-6 52.00 EUR
- 301 Alexander Engl: CHEMIE PUR – Unterrichten in der Natur: *Entwicklung und Evaluation eines kontextorientierten Unterrichtskonzepts im Bereich Outdoor Education zur Veränderung der Einstellung zu „Chemie und Natur“*
ISBN 978-3-8325-5174-2 59.00 EUR
- 302 Christin Marie Sajons: Kognitive und motivationale Dynamik in Schülerlaboren. *Kontextualisierung, Problemorientierung und Autonomieunterstützung der didaktischen Struktur analysieren und weiterentwickeln*
ISBN 978-3-8325-5155-1 56.00 EUR
- 303 Philipp Bitzenbauer: Quantenoptik an Schulen. *Studie im Mixed-Methods Design zur Evaluation des Erlanger Unterrichtskonzepts zur Quantenoptik*
ISBN 978-3-8325-5123-0 59.00 EUR
- 304 Malte S. Ubben: Typisierung des Verständnisses mentaler Modelle mittels empirischer Datenerhebung am Beispiel der Quantenphysik
ISBN 978-3-8325-5181-0 43.50 EUR
- 305 Wiebke Kuske-Janßen: Sprachlicher Umgang mit Formeln von LehrerInnen im Physikunterricht am Beispiel des elektrischen Widerstandes in Klassenstufe 8
ISBN 978-3-8325-5183-4 47.50 EUR

- 306 Kai Bliesmer: Physik der Küste für außerschulische Lernorte. *Eine Didaktische Rekonstruktion*
ISBN 978-3-8325-5190-2 58.00 EUR
- 307 Nikola Schild: Eignung von domänenspezifischen Studieneingangsvariablen als Prädiktoren für Studienerfolg im Fach und Lehramt Physik
ISBN 978-3-8325-5226-8 42.00 EUR
- 308 Daniel Averbeck: Zum Studienerfolg in der Studieneingangsphase des Chemiestudiums. *Der Einfluss kognitiver und affektiv-motivationaler Variablen*
ISBN 978-3-8325-5227-5 51.00 EUR
- 309 Martina Strübe: Modelle und Experimente im Chemieunterricht. *Eine Videostudie zum fachspezifischen Lehrerwissen und -handeln*
ISBN 978-3-8325-5245-9 45.50 EUR
- 310 Wolfgang Becker: Auswirkungen unterschiedlicher experimenteller Repräsentationen auf den Kenntnisstand bei Grundschulkindern
ISBN 978-3-8325-5255-8 50.00 EUR
- 311 Marvin Rost: Modelle als Mittel der Erkenntnisgewinnung im Chemieunterricht der Sekundarstufe I. *Entwicklung und quantitative Dimensionalitätsanalyse eines Testinstruments aus epistemologischer Perspektive*
ISBN 978-3-8325-5256-5 44.00 EUR
- 312 Christina Kobl: Förderung und Erfassung der Reflexionskompetenz im Fach Chemie
ISBN 978-3-8325-5259-6 41.00 EUR

Alle erschienenen Bücher können unter der angegebenen ISBN direkt online (<http://www.logos-verlag.de>) oder per Fax (030 - 42 85 10 92) beim Logos Verlag Berlin bestellt werden.

Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf, Hans Niedderer, Mathias Ropohl und Elke Sumfleth

Die Reihe umfasst inzwischen eine große Zahl von wissenschaftlichen Arbeiten aus vielen Arbeitsgruppen der Physik- und Chemiedidaktik und zeichnet damit ein gültiges Bild der empirischen physik- und chemiedidaktischen Forschung im deutschsprachigen Raum.

Die Herausgeber laden daher Interessenten zu neuen Beiträgen ein und bitten sie, sich im Bedarfsfall an den Logos-Verlag oder an ein Mitglied des Herausgeberteams zu wenden.

Kontaktadressen:

Univ.-Prof. Dr. Martin Hopf
Universität Wien,
Österreichisches Kompetenzzentrum
für Didaktik der Physik,
Porzellangasse 4, Stiege 2,
1090 Wien, Österreich,
Tel. +43-1-4277-60330,
e-mail: martin.hopf@univie.ac.at

Prof. Dr. Hans Niedderer
Inst. f. Didaktik der Naturwissenschaften,
Abt. Physikdidaktik,
FB Physik/ Elektrotechnik,
Universität Bremen,
Postfach 33 04 40, 28334 Bremen
Tel. 0421-218 4695 (Sekretariat),
e-mail: niedderer@physik.uni-bremen.de

Prof. Dr. Mathias Ropohl
Didaktik der Chemie,
Fakultät für Chemie,
Universität Duisburg-Essen,
Schützenbahn 70, 45127 Essen,
Tel. 0201-183 2704,
e-mail: mathias.ropohl@uni-due.de

Prof. Dr. Elke Sumfleth
Didaktik der Chemie,
Fakultät für Chemie,
Universität Duisburg-Essen,
Schützenbahn 70, 45127 Essen
Tel. 0201-183 3757/3761,
e-mail: elke.sumfleth@uni-due.de

Das Nutzen von Modellen ist konstitutiv für die Fachwissenschaft der Chemie und in der Folge auch für chemiebezogene Lehr-/Lernumgebungen. Um die Modellnutzung im schulischen Chemieunterricht erfolgreich zu integrieren, bedarf es daher der systematischen Aufbereitung des vielschichtigen Begriffs „Modell“, der Verknüpfung mit kompetenzorientierten Bildungsstandards sowie der Prüfung der theoretischen Konzeptualisierung an empirischen Befunden.

In der vorliegenden Arbeit wurde zu diesem Zweck ein chemiespezifischer Begriffsvorschlag gemacht, der erkenntnistheoretische und fachdidaktische Arbeiten in einem messbaren Strukturvorschlag für die hypothetisch-deduktive Modellnutzung synthetisiert. Die so erhaltene Beschreibung von Modellobjekten mit vier hierarchisch geschachtelten Modellkomponenten (kleinste Sinneinheiten, Relationen, Operationen, Regeln) wurde für problemorientierte Multiple-Choice-Leistungsaufgaben operationalisiert. Nach mehreren qualitativen und quantitativen Iterationen zur Qualitätssicherung des Testinstruments blieb diese Vierdimensionalität nach Item-Response-Theorie-Skalierungen empirisch plausibel und kann zukünftig zur Orientierung in einem sehr heterogenen Begriffsfeld dienen.

Die Studie zeigt, dass die vorgeschlagene Strukturierung von Modellnutzung in der Chemie zielführend ist und trägt auf diese Weise zu einer übersichtlicheren Anwendung beim Lehren und Lernen von Chemie bei.

Logos Verlag Berlin

ISBN 978-3-8325-5256-5