

**Judith Pollmeier**

**Kontextmerkmale und die Bearbeitung von Aufgaben in einem Test  
naturwissenschaftlicher Kompetenz in der Grundschule**

**2015**

Fach Psychologie

Dissertationsthema

**Kontextmerkmale und die Bearbeitung von Aufgaben in einem Test  
naturwissenschaftlicher Kompetenz in der Grundschule**

Inaugural-Dissertation zur Erlangung des Doktorgrades im Fachbereich Psychologie und  
Sportwissenschaft der Westfälischen Wilhelms-Universität in Münster

Vorgelegt von  
*Judith Pollmeier*  
aus Ahlen (Westf.)  
2015

---

Dekanin/Dekan:	Prof. Dr. Manfred Holodynski
Erste Gutachterin/Erster Gutachter:	Prof. Dr. Kornelia Möller
Zweite Gutachterin/Zweiter Gutachter:	Prof. Dr. Manfred Holodynski
Tag der mündlichen Prüfung(en):	02.07.2015
Tag der Promotion:	02.07.2015

## **Danksagung**

Die vorliegende Arbeit nutzt Daten, welche im Rahmen der Querschnittstudie des Projekts Science-P (Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule) erhoben wurden. Das Projekt wurde von der Deutschen Forschungsgemeinschaft im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ gefördert. Ich danke daher Prof. Dr. Kornelia Möller, Prof. Dr. Beate Sodian, Prof. Dr. Susanne Koerber, Prof. Dr. Ilonca Hardy, Prof. Dr. Knut Schwippert, Dr. Nicola Kropf, Dr. Daniela Mayer, Dr. Thilo Kleickmann und Steffen Tröbst, welche alle in vielfältiger Weise bis zum Zeitpunkt der Querschnittstudie an der Durchführung des Projektes Science-P beteiligt waren. Ohne die Mitarbeit zahlreicher studentischer Hilfskräfte wären die aufwändigen Datenerhebungen der beiden vorliegenden Studien nicht möglich gewesen. Darüber hinaus gilt mein besonderer Dank Sina Marten, welche im Rahmen ihrer Masterarbeit die Daten zur Komponente Verdunstung für die Experimentalstudie erhoben hat. Ich bedanke mich bei Prof. Dr. Kornelia Möller und Prof. Dr. Manfred Holodynski für ihre Anmerkungen zu einer früheren Version dieser Arbeit. Abschließend danke ich Prof. Dr. Rainer Watermann für seine Bereitschaft als drittes Mitglied meines Promotionskomitees zu fungieren.

## Inhaltsverzeichnis

Verzeichnis der Abkürzungen.....	9
Verzeichnis der Tabellen.....	10
Verzeichnis der Abbildungen.....	12
Zusammenfassung.....	13
1    Einleitung.....	14
2    Theoretischer Hintergrund und Forschungsstand.....	17
2.1    Kompetenzen, Kompetenzmodelle und Kompetenzentwicklung.....	17
2.1.1    Begriffliche Klärung, Grundlagen und Funktionen der Kompetenzmodellierung.....	17
2.1.2    Kompetenzmodelle in den Naturwissenschaften.....	19
2.1.2.1    Normative Kompetenzmodelle in den Naturwissenschaften.....	19
2.1.2.2    Deskriptive Kompetenzmodelle in den Naturwissenschaften.....	21
2.1.2.3    Naturwissenschaftliche Kompetenzmodelle mit theoretischer Begründung und empirischer Absicherung.....	22
2.1.2.4    Das Projekt Science-P.....	24
2.1.2.5    Zusammenfassung und Fazit.....	25
2.2    Schülervorstellungen in den Naturwissenschaften.....	27
2.2.1    Schülervorstellungen und Conceptual Change.....	27
2.2.2    Kohärenz vs. Fragmentierung?.....	30
2.2.3    Die Kontextualität von Schülervorstellungen.....	34
2.2.4    Untersuchungen zu Schülervorstellungen bei Verdunstung und Kondensation.....	36
2.2.4.1    Schülervorstellungen zu Verdunstung und Kondensation.....	36
2.2.4.2    Kontextualität von Schülervorstellungen zu Verdunstung und Kondensation.....	38
2.2.5    Untersuchungen zu Schülervorstellungen zum Schwimmen und Sinken.....	39
2.2.5.1    Schülervorstellungen zum Schwimmen und Sinken.....	39
2.2.5.2    Kontextualität von Schülervorstellungen zum Schwimmen und Sinken.....	41
2.2.6    Conceptual Change, Kompetenzmodelle und Kontexteinfluss.....	42
2.3    Aufgabenmerkmale.....	43
2.3.1    Die Rolle von Aufgabenmerkmalen bei der Validierung von Kompetenzmodellen.....	43
2.3.2    Unterschiedliche Arten von Aufgabenmerkmalen.....	45
2.3.2.1    Systematisierung von Aufgabenmerkmalen.....	45
2.3.2.2    Formale Aufgabenmerkmale.....	45
2.3.2.3    Inhaltliche Aufgabenmerkmale.....	46
2.3.2.4    Prozessorientierte Aufgabenmerkmale.....	46
2.3.3    Merkmale des Kontexts.....	47
2.3.3.1    Kontextmerkmale und Erfassung von Schülervorstellungen.....	47
2.3.3.2    Relevante Kontextmerkmale bei Verdunstung und Kondensation.....	48
2.3.3.3    Relevante Kontextmerkmale bei Schwimmen und Sinken.....	50

3	Zielsetzung, Fragestellungen und Hypothesen.....	52
4	Studie 1: Querschnittstudie .....	57
4.1	Einbettung der Studie 1 in das Projekt Science-P .....	57
4.2	Präzisierung und Einschränkung der Fragestellungen und Hypothesen für Studie 1.....	57
4.3	Methode Studie 1 .....	59
4.3.1	Anlage und Aufbau der Untersuchung in Studie 1.....	59
4.3.1.1	Instrumente zur Erfassung des Naturwissenschaftlichen Wissens in Studie 1 .....	59
4.3.1.2	Weitere Erhebungsinstrumente in Studie 1 .....	62
4.3.1.3	Stichprobe der Studie 1 .....	62
4.3.1.4	Durchführung der Studie 1 .....	64
4.3.2	Datenaufbereitung und –auswertung in Studie 1.....	64
4.3.2.1	Kodierung der Schülerantworten auf die Aufgaben in Studie 1.....	64
4.3.2.2	Kodierung der Aufgaben hinsichtlich der untersuchten Merkmale in Studie 1 .....	65
4.3.2.3	Dimensionalitätsprüfung, Itemselektion und zentrale psychometrische Kennwerte der Skalen und Items in Studie 1 .....	66
4.3.2.4	Vorgehen und Verfahren zur Hypothesenprüfung in Studie 1: Generalisierte lineare gemischte Modelle mit gekreuzten zufälligen Effekten .....	67
4.4	Ergebnisse von Studie 1 .....	71
4.4.1	Ergebnisse der Aufgabenanalyse von Studie 1 .....	71
4.4.1.1	Verdunstung und Kondensation .....	71
4.4.1.2	Schwimmen und Sinken.....	74
4.4.2	Deskriptive Ergebnisse von Studie 1.....	77
4.4.2.1	Verdunstung und Kondensation .....	77
4.4.2.2	Schwimmen und Sinken.....	78
4.4.3	Ergebnisse der inferenzstatistischen Analysen von Studie 1 .....	78
4.4.3.1	Verdunstung und Kondensation .....	78
4.4.3.2	Schwimmen und Sinken.....	83
4.5	Diskussion von Studie 1 .....	87
4.5.1	Zusammenfassung und Diskussion der Ergebnisse von Studie 1 .....	87
4.5.1.1	F1: Vorliegen der Kontextmerkmale in den Aufgaben von Studie 1 .....	87
4.5.1.2	F2: Effekt der Kontextmerkmale auf die Aufgabenschwierigkeiten in Studie 1 .....	90
4.5.1.3	F3: Interaktion zwischen Kontextmerkmalen und Kompetenzniveaus.....	92
4.5.1.4	F4: Effekt der formalen Merkmale auf die Aufgabenschwierigkeiten in Studie 1....	93
4.5.1.5	Zusammenfassung.....	94
4.5.2	Beschränkungen der Studie 1 .....	95
4.5.3	Konsequenzen für die Aufgabenentwicklung aus Studie 1 .....	96
4.5.4	Fazit aus Studie 1 und Ausblick .....	97
5	Studie 2: Experimentelle Variation von Aufgabenmerkmalen .....	99
5.1	Einbettung der Studie 2 in das Projekt Science-P .....	99

5.2	Präzisierung und Einschränkung der Fragestellungen und Hypothesen für Studie 2.....	99
5.3	Methode Studie 2 .....	101
5.3.1	Anlage und Aufbau der Untersuchung in Studie 2.....	101
5.3.1.1	Stichprobe der Studie 2 .....	101
5.3.1.2	Instrumente zur Erfassung des Naturwissenschaftlichen Wissens in Studie 2.....	102
5.3.1.3	Weitere Erhebungsinstrumente in Studie 2 .....	106
5.3.1.4	Durchführung der Studie 2 .....	106
5.3.2	Datenaufbereitung und –auswertung in Studie 2.....	107
5.3.2.1	Kodierung der Schülerantworten auf die Aufgaben in Studie 2.....	107
5.3.2.2	Kodierung der Aufgaben hinsichtlich der untersuchten Merkmale in Studie 2 .....	107
5.3.2.3	Dimensionalitätsprüfung, Itemselektion und zentrale psychometrische Kennwerte der Skalen und Items in Studie 2.....	107
5.3.2.4	Vorgehen und Verfahren zur Hypothesenprüfung in Studie 2: Generalisierte lineare gemischte Modelle mit gekreuzten zufälligen Effekten .....	109
5.4	Ergebnisse von Studie 2 .....	110
5.4.1	Deskriptive Ergebnisse von Studie 2.....	110
5.4.1.1	Verdunstung .....	110
5.4.1.2	Kondensation.....	111
5.4.1.3	Dichte und Verdrängung .....	112
5.4.2	Ergebnisse der inferenzstatistischen Analysen von Studie 2 .....	113
5.4.2.1	Verdunstung .....	113
5.4.2.2	Kondensation.....	115
5.4.2.3	Dichte und Verdrängung .....	116
5.5	Diskussion von Studie 2.....	117
5.5.1	Zusammenfassung und Diskussion der Ergebnisse von Studie 2 .....	117
5.5.1.1	F2: Effekt der Kontextmerkmale auf die Aufgabenschwierigkeiten in Studie 2 .....	118
5.5.1.2	F3: Interaktion zwischen Kontextmerkmalen und Kompetenzniveaus .....	120
5.5.1.3	FZ: Einfluss des konkreten Iteminhalts auf die Aufgabenschwierigkeiten in Studie 2 .....	120
5.5.1.4	Zusammenfassung.....	121
5.5.2	Beschränkungen der Studie 2 .....	121
5.5.3	Konsequenzen für die Aufgabenentwicklung aus Studie 2 .....	122
5.5.4	Fazit aus Studie 2 und Ausblick .....	123
6	Fazit.....	125
6.1	Zusammenfassung und Diskussion der beiden Studien .....	125
6.2	Konsequenzen für die Aufgabenentwicklung .....	127
6.3	Konsequenzen für die Kompetenzmodellierung .....	128
6.4	Bedeutung der Ergebnisse der Studien 1 und 2 für die Debatte um Kohärenz vs. Fragmentierung .....	129

6.5	Ausblick .....	132
7	Literatur.....	135
8	Anhang .....	142
9	Verzeichnis der elektronischen Anhänge .....	167
	Plagiatserklärung.....	168
	Lebenslauf.....	169



## Verzeichnis der Abkürzungen

A	Komponente Auftrieb
D	Komponente Dichte
FC	Aufgabe mit dem Format <i>forced choice</i>
K	Komponente Kondensation
MC	Aufgabe mit dem Format <i>multiple choice</i>
MS	Aufgabe mit dem Format <i>multiple select</i>
NV	Kompetenzniveau Naive Vorstellungen
Science-P	DFG-Projekt „Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule“
SL	<i>scientific literacy</i> (naturwissenschaftliche Grundbildung)
SuS	Themenbereich Schwimmen/Sinken
V	Komponente Verdrängung
Vd	Komponente Verdunstung
VuK	Themenbereich Verdunstung/Kondensation
WV	Kompetenzniveau Wissenschaftliche Vorstellungen
ZV	Kompetenzniveau Zwischenvorstellungen

## Verzeichnis der Tabellen

<b>Tabelle 1</b> Untersuchte Aufgabenmerkmale und deren angenommen Wirkrichtung auf die Aufgabenschwierigkeit .....	54
<b>Tabelle 2</b> Übersicht über eingesetzte Aufgaben und Items in der Querschnittstudie .....	62
<b>Tabelle 3</b> Stichprobenbeschreibung .....	63
<b>Tabelle 4</b> Übersicht über inferenzstatistische Auswertungen in Studie 1 für den Themenbereich Verdunstung/Kondensation .....	70
<b>Tabelle 5</b> Übersicht über inferenzstatistische Auswertungen in Studie 1 für den Themenbereich Schwimmen/Sinken .....	71
<b>Tabelle 6</b> Verteilung der Items zu Verdunstung und Kondensation auf die Aufgabenmerkmale, Itemmittelwerte und Standardabweichungen .....	71
<b>Tabelle 7</b> Zusammenhänge zwischen den untersuchten Merkmalen des Kontexts und dem Kompetenzniveau für den Themenbereich Verdunstung und Kondensation .....	73
<b>Tabelle 8</b> Verteilung der Items zu Schwimmen und Sinken auf die Aufgabenmerkmale, Itemmittelwerte und Standardabweichungen .....	75
<b>Tabelle 9</b> Zusammenhänge zwischen den untersuchten Merkmalen des Kontexts und dem Kompetenzniveau für den Themenbereich Schwimmen und Sinken .....	76
<b>Tabelle 10</b> Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Verdunstung und Kondensation .....	79
<b>Tabelle 11</b> Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Verdunstung .....	80
<b>Tabelle 12</b> Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Kondensation .....	81
<b>Tabelle 13</b> Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Schwimmen und Sinken .....	83
<b>Tabelle 14</b> Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Auftrieb .....	84
<b>Tabelle 15</b> Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Dichte .....	85
<b>Tabelle 16</b> Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Verdrängung .....	85
<b>Tabelle 17</b> Stichprobenbeschreibung Studie 2 .....	101
<b>Tabelle 18</b> Verteilung der Aufgaben zu Verdunstung auf die Kontextmerkmale in Studie 2 .....	103
<b>Tabelle 19</b> Verteilung der Aufgaben zu Kondensation auf die Kontextmerkmale in Studie 2 .....	104

<b>Tabelle 20</b> Verteilung der Aufgaben zu Dichte und Verdrängung auf das Kontextmerkmal in Studie 2 .....	106
<b>Tabelle 21</b> Deskriptive Statistiken der Items zu Verdunstung in Studie 2 .....	110
<b>Tabelle 22</b> Deskriptive Statistiken der Items zu Kondensation in Studie 2.....	111
<b>Tabelle 23</b> Deskriptive Statistiken der Items zu Dichte und Verdrängung in Studie 2 .....	112
<b>Tabelle 24</b> Ergebnisse der Prüfung des Einflusses der Aufgabenmerkmale für Verdunstung .....	113
<b>Tabelle 25</b> Ergebnisse der Prüfung des Einflusses der Aufgabenmerkmale für Kondensation ....	115
<b>Tabelle 26</b> Ergebnisse der Prüfung des Einflusses der Aufgabenmerkmale für Dichte .....	116
<b>Tabelle 27</b> Ergebnisse der Prüfung des Einflusses der Aufgabenmerkmale für Verdrängung.....	117

## Verzeichnis der Abbildungen

<b>Abbildung 1</b> Ankeraufgabe zur Komponente Verdunstung im Antwortformat <i>multiple select</i> ....	61
<b>Abbildung 2</b> Mittlere geschätzte Schwierigkeiten der Items zu Verdunstung getrennt nach Kompetenzniveau und Sichtbarkeit des Wasserspiegels .....	82
<b>Abbildung 3</b> Mittlere geschätzte Schwierigkeiten der Items zu Auftrieb getrennt nach Kompetenzniveau und Größe des Gegenstandes .....	87
<b>Abbildung 4</b> Ankeraufgabe zum Bereich Dichte, Aufgabentyp inkongruent .....	105
<b>Abbildung 5</b> Mittlere geschätzte Schwierigkeiten der Items zu Verdunstung getrennt nach Inhalt .....	114
<b>Abbildung 6</b> Mittlere geschätzte Schwierigkeiten der Items zu Kondensation getrennt nach Inhalt .....	116

## **Zusammenfassung**

Die Beschreibung von Kompetenzen und ihrer Entwicklung mittels theoretisch fundierter und empirisch abgesicherter Modelle ist von großer Bedeutung für die Überprüfung von Bildungszielen. Im Bereich der Entwicklung naturwissenschaftlicher Kompetenzen besteht derzeit allerdings noch ein Mangel an entsprechenden Modellen in Bezug auf die Grundschule. Dabei stellt in dieser Altersstufe insbesondere die Entwicklung von validen Testinstrumenten zur Erfassung dieser Kompetenzen eine große Herausforderung dar. Ein Rückgriff auf die Erkenntnisse der Forschung zu vorunterrichtlichen Schülervorstellungen sowie deren Veränderungen (Conceptual Change) kann zur Formulierung gültiger Instrumente und Modelle beitragen.

Die Theorie des Conceptual Change ist ein zentrales Paradigma, auf welchem naturwissenschaftliche Fachdidaktik, besonders im Grundschulbereich, Forschungs- und Lehransätze begründet. Innerhalb der Conceptual Change-Forschung werden verschiedene offene Forschungsfragen diskutiert. Der Kontext der Anwendung von Schülervorstellungen nimmt dabei eine wesentliche Rolle ein: Verschiedene Studien konnten eine Abhängigkeit geäußerter Schülervorstellungen von dem jeweiligen Kontext belegen.

Obwohl auch für die Kompetenzmodellierung der Kontext per Definition ein wichtiger Aspekt ist, wird dieser in neueren Ansätzen zur Modellierung naturwissenschaftlicher Kompetenz in der Grundschule nicht umfänglich berücksichtigt. Meine Arbeit geht daher der Frage nach, ob kontextuelle Merkmale von schriftlichen, geschlossenen Aufgaben zur Erfassung naturwissenschaftlicher Kompetenz (in den Themenbereichen Schwimmen/Sinken und Verdunstung/Kondensation) das Antwortverhalten von Grundschulkindern beeinflussen.

Eine Analyse der Daten von 1820 Schülerinnen und Schülern der zweiten, dritten und vierten Klasse führte allerdings zu einem uneinheitlichen Bild: Es wurden zwar Hinweise auf eine Kontextabhängigkeit naturwissenschaftlicher Kompetenz gefunden. Diese ließen sich aber nicht auf die Mehrheit der untersuchten Kontextmerkmale oder beide untersuchten Themenbereiche generalisieren. Jedoch waren die Items der Studien nicht mit der Intention der Untersuchung von Aufgabenmerkmalen entwickelt worden.

Daher wurden in einer zweiten, experimentell angelegten Studie die Aufgaben der ersten Studie einer Überarbeitung unterzogen und hinsichtlich der schon untersuchten Merkmale systematisch variiert und erneut mit den gleichen Methoden analysiert. Es zeigte sich, dass vor allem in den Bereichen Dichte und Verdrängung als Facetten des Themenbereichs Schwimmen/Sinken das Kontextmerkmal Kongruenz einen massiven Teil der Varianz aufklären konnte, während in den Facetten Verdunstung und Kondensation der konkrete Inhalt der Items der erklärungsmächtigste Faktor war. Eine Kontextualisierung von Schülervorstellungen scheint demnach differenziell bei unterschiedlichen Themenbereichen aufzutreten, während nichtsdestotrotz Erklärungsansätze unterschiedlicher Niveaus parallel auf verschiedene Kontexte angewendet werden.

## 1 Einleitung

Im Zuge der Definition von Bildungszielen für verschiedene Fächer und Schulstufen wird die Beschreibung von Kompetenzen mittels Kompetenzmodellen gefordert (Schecker & Parchmann, 2006). Solchen Modellen kommt eine Schlüsselfunktion für die Entwicklung von Messinstrumenten, die Diagnostik von Kompetenzen und die optimale Steuerung von Bildungssystemen zu (Klieme & Leutner, 2006). Für den Bereich der Naturwissenschaften wurden bereits einige Modelle entwickelt, die entweder theoretisch oder empirisch basiert die Struktur, verschiedene Ausprägungen oder die Entwicklung von Kompetenzen beschreiben (vgl. Beinbrech & Möller, 2008; Schecker & Parchmann, 2006). Besonders die Modellierung des Erwerbs und der Entwicklung von Kompetenzen erscheint hinsichtlich einer fachdidaktischen Nutzung von Kompetenzmodellen erstrebenswert (Beinbrech & Möller, 2008). Der Bereich naturwissenschaftlicher Kompetenzen in der Grundschule ist in existierenden Modellen noch unterrepräsentiert.

Um ein valides Modell abzuleiten und aussagekräftige Messinstrumente zu entwickeln, kann der Rückgriff auf bereits bestehende Forschung zu naturwissenschaftlichen Kompetenzen in der Grundschule wichtige Hinweise liefern. Ein zentrales Paradigma, in welchem das naturwissenschaftliche Wissen von Grundschulern erforscht wurde, stellt die Theorie des Conceptual Change dar (z.B. Posner, Strike, Hewson & Gertzog, 1982), deren zentrale Annahme die Veränderung von vorunterrichtlichen Schülervorstellungen zu naturwissenschaftlichen Phänomenen durch konstruktive Lernprozesse ist. Die Erforschung dieser Schülervorstellungen und ihres Wandels blickt mittlerweile auf eine jahrzehntelange Tradition zurück (Carey, 1985). Dennoch blieben einige Fragen bislang unbeantwortet. Insbesondere im Hinblick auf die Struktur des initialen Wissens stehen sich bislang zwei, scheinbar unvereinbare, Ansätze gegenüber: der Kohärenz- (Vosniadou, 1994; Vosniadou, Vamvakoussi & Skopeliti, 2008) und der Fragmentierungsansatz (diSessa, 2006; 2008). In Bezug auf das Anliegen der Entwicklung eines möglichst validen Kompetenzmodells und dazugehörigen Messinstruments ist besonders der Einfluss des Kontexts auf Schülervorstellungen relevant. Der Kontext ist ein kritisches Element der Vorstellungen und häufig werden diese kontextspezifisch geäußert.

Auch für die Erhebung von naturwissenschaftlichen Kompetenzen ist die Frage nach dem Kontext relevant: Die Erfassung von Kompetenzen soll in problemhaltigen Situationen (also Anwendungskontexten) verankert sein (Klieme & Leutner, 2006; Weinert, 2002). Merkmale dieser Problemsituationen können analysiert und zur Beschreibung von Aufgaben genutzt werden. Darüber hinaus bietet sich die Möglichkeit, sie auf ihr Wirken auf das Antwortverhalten hin zu untersuchen und die gewonnen Erkenntnisse für die Validierung von Kompetenzmodellen zu nutzen (z.B. Hartig & Frey, 2012). Eine Beschreibung der Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule könnte also auf relevanten Erkenntnissen der Conceptual Change-

Forschung zum Kontexteinfluss aufbauen und helfen die noch nicht abschließend bearbeitete Frage der Kontextualität von Schülervorstellungen zu beantworten.

Zentrales Anliegen meiner Arbeit ist daher die Untersuchung von Aufgabenmerkmalen, speziell Merkmalen des Aufgabenkontexts, in einem Test zur Erfassung naturwissenschaftlicher Kompetenz von Grundschulern in den beiden Themenbereichen Schwimmen/Sinken und Verdunstung/Kondensation. Die Ergebnisse einer solchen Untersuchung können potentiell Hinweise für die Formulierung von Modellen der Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule liefern und die Validität zugehöriger Testmodelle steigern. Ein möglicher Nebenertrag der Untersuchung dieser Fragestellung sind Hinweise in Bezug auf die Diskussion um Kohärenz oder Fragmentierung initialer Schülervorstellungen. Die Arbeit gliedert sich grob in einen theoretischen und einen empirischen Teil.

Im ersten Segment des theoretischen Teils der Arbeit werden zunächst die Grundlagen der Kompetenzmodellierung erörtert (Kapitel 2.1.1), bevor in Kapitel 2.1.2 verschiedene Ansätze zur Modellierung naturwissenschaftlicher Kompetenz vorgestellt werden. Die verschiedenen Ansätze werden daraufhin zusammenfassend verglichen und dahingehend analysiert, welche Aspekte der Conceptual Change-Forschung in den Modellen berücksichtigt werden und welche noch nicht ausreichend beachtet werden (Kapitel 2.1.2.5).

Der aktuelle Forschungsstand zu Schülervorstellungen und Conceptual Change ist der zweite theoretische Schwerpunkt der Arbeit in Kapitel 2.2. Dabei wird ein Augenmerk auf die Darstellung der Ansätze von Kohärenz und Fragmentierung (Kapitel 2.2.2) sowie insbesondere die Bedeutung des Kontextes für Schülervorstellungen gelegt (Kapitel 2.2.3). Daraufgehend werden Untersuchungen zu Schülervorstellungen in den beiden betrachteten Themenbereichen vorgestellt und gesondert nach Hinweise auf potentiell relevante Kontextmerkmale evaluiert (Kapitel 2.2.4 und 2.2.5).

Der letzte theoretische Fokus wird auf die Bedeutung und Klassifizierung von Aufgabenmerkmalen gelegt (Kapitel 2.3). Erneut wird hierbei ebenfalls die gesonderte Rolle von Kontextmerkmalen berücksichtigt (Kapitel 2.3.3).

Ausgehend von den theoretischen Erörterungen werden in Kapitel 3 Fragestellungen und Hypothesen abgeleitet, die anschließend in zwei empirischen Studien untersucht werden.

Bei der ersten empirischen Studie handelt es sich um eine querschnittliche Untersuchung von Grundschulern der zweiten, dritten und vierten Jahrgangsstufe (Kapitel 4). Das zu diesem Zweck eingesetzte Aufgabenmaterial wurde nicht zur Analyse des Einflusses von Kontextmerkmalen entwickelt. Daher werden die eingesetzten Items zunächst deskriptiv hinsichtlich verschiedener Merkmale betrachtet (Kapitel 4.4.1). Anschließend wird mittels linear-gemischter Modelle eine inferenzstatistische Überprüfung des Einflusses der Merkmale auf die empirischen Itemsschwierigkeiten vorgenommen (Kapitel 4.4.3). Die Ergebnisse der Analysen werden

anschließend in Kapitel 4.5 diskutiert und die gesamte Studie hinsichtlich methodischer Einschränkungen besprochen (Kapitel 4.5.2).

Die Behebung dieser methodischen Mängel war das Ziel der anschließend beschriebenen experimentellen Studie (Kapitel 5). Hierfür wurden Kontextmerkmale von Aufgaben systematisch variiert (Kapitel 5.3.1) und ihre Auswirkungen anschließend analog zu ersten Studie analysiert (Kapitel 5.4). Die Ergebnisse werden in Kapitel 5.5 diskutiert.

Den Abschluss meiner Arbeit bildet eine Zusammenfassung und Diskussion der Ergebnisse beider Studien in Kapitel 6. Diese werden vor allem hinsichtlich ihrer Bedeutung für die Aufgabenentwicklung (Kapitel 6.2) und die Kompetenzmodellierung (Kapitel 6.3) besprochen, bevor die Bedeutung der Ergebnisse für die Ansätze von Kohärenz oder Fragmentierung diskutiert wird (Kapitel 6.4). Abschließend werden einige offene Forschungsfragen und lohnenswerte Ansätze für weitere Untersuchungen skizziert (Kapitel 6.5).



## **2 Theoretischer Hintergrund und Forschungsstand**

### **2.1 Kompetenzen, Kompetenzmodelle und Kompetenzentwicklung**

#### **2.1.1 Begriffliche Klärung, Grundlagen und Funktionen der Kompetenzmodellierung**

Nach dem unerwartet schlechten Abschneiden deutscher Schüler in internationalen Vergleichsstudien wie TIMSS (Baumert, Bos & Lehmann, 2000) oder PISA (Baumert et al., 2001), hat sich in Deutschland eine Diskussion um die Qualität des Schulsystems entwickelt. Diese hat vor allem eine Änderung des Denkens weg von einer Input-Orientierung hin zu einer Output-Orientierung bewirkt, bei der die Ergebnisse des Schulsystems, also die Lernergebnisse, mit festgelegten Kriterien verglichen und bewertet werden sollen (Klieme et al., 2003). Diese Kriterien sind als Kompetenzerwartungen oder offiziell in sog. Bildungsstandards festgehalten worden (Klieme et al., 2003). Für den Abgleich mit den Kompetenzerwartungen ist eine Messung der Schülerleistungen notwendig. In diesem Zusammenhang erfüllen Kompetenzmodelle den Zweck, die Lücke zwischen Anforderungsbeschreibungen und Testaufgaben zur Leistungsmessung zu schließen (Schecker & Parchmann, 2006).

Die in solchen Modellen beschriebenen Kompetenzen werden dabei verstanden als

„die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert 2002, S. 27 f.).

Der in dieser Definition beschriebene, direkte Bezug auf die Anforderungen in bestimmten Situationen ist dabei ein wesentliches Charakteristikum, das den Kompetenzbegriff vor allem vom Intelligenzbegriff abgrenzt. In der Intelligenzforschung werden allgemeine, kontextunabhängige Fähigkeiten untersucht (Klieme & Leutner, 2006; McClelland, 1973; Weinert, 2001). Die Bezugnahme auf spezifische Situationen erleichtert einerseits die Definition untersuchter Kompetenzen, da ausgehend von den jeweiligen Anforderungen beschrieben werden kann, welche Fähigkeiten zur Problemlösung notwendig sind. Andererseits haben die so definierten Kompetenzen eine klare Funktionalität, die bei stärkerer Abstraktion der untersuchten Fähigkeiten verloren gehen kann (McClelland, 1973; Weinert, 2001). Allerdings liegt gerade in dieser Kontextabhängigkeit auch eine besondere Schwierigkeit für die Entwicklung von Messinstrumenten, da eine systematische Variation von Kontexten für die Messung nicht oder nur sehr begrenzt möglich ist (Weinert, 2001).

Weitere Abgrenzung zum Intelligenzbegriff erfährt diese Definition außerdem durch die Ausweitung von rein kognitiven auf motivationale, volitionale und soziale Komponenten, denen eine entscheidende Rolle bei der tatsächlich gezeigten Leistung zugeschrieben wird (Weinert,

2001), sowie der Erlernbarkeit von Kompetenzen. Der letzte Punkt soll vor allem der möglichen Diskriminierung von Personen oder Gruppen entgegenwirken, die im Zusammenhang mit dem Intelligenzquotienten kritisiert wurde (McClelland, 1973; Weinert, 2001), und kann außerdem als Anhaltspunkt zur Beurteilung der Validität von Messinstrumenten dienen, da Lernende, die an einem Programm zur Steigerung der in Frage stehenden Kompetenz teilnehmen, auch tatsächlich einen Zuwachs dieser Kompetenz zeigen sollten (Weinert, 2001). Deutlich abzugrenzen ist der Kompetenzbegriff laut Weinert (2001) außerdem von Schlüsselkompetenzen, die in Zusammenhang mit lebenslangem Lernen diskutiert werden, und Metakompetenzen, also dem deklarativen oder prozeduralen Wissen über die eigenen Kompetenzen.

In der vorliegenden Arbeit werde ich mich auf einen konkreten Ausschnitt aus diesem aufgespannten Spektrum beziehen, nämlich auf „kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen“ (Klieme & Leutner, 2006, S. 879). Im Vergleich zu der breiten Definition von Weinert (2002) sollen also motivationale, volitionale und soziale Aspekte hier nicht betrachtet werden. Darüber hinaus wird in dieser verkürzten Definition der Bezug auf ein konkretes Fachgebiet, für welches die untersuchten Kompetenzen relevant sind, herausgestellt. Die Kontextspezifizität wird in dieser Definition ebenfalls noch einmal stärker expliziert, und so fordern die Autoren auch den Einbezug entsprechender Merkmale von Anforderungssituationen in die psychometrische Modellierung von Kompetenzen, um Testwerte somit direkt auf bewältigte Anforderungen beziehen zu können (Klieme & Leutner, 2006). Damit nehmen die Merkmale der als Aufgaben umgesetzten Anforderungssituationen außerdem eine wichtige Stellung bei der Konstruktvalidierung von Kompetenzen und den zugehörigen Modellen ein (Hartig & Frey, 2012; Whitely, 1983; siehe Kapitel 2.3.1). Desweiteren liefern sie wichtige Informationen für die kriteriumsorientierte Skalenbeschreibung und die weitere Itementwicklung (Hartig & Frey, 2012). Eine genaue Operationalisierung von Kompetenzen wird in Kompetenzmodellen vorgenommen, die sowohl eine Beschreibung der Struktur bzw. Dimensionalität von Kompetenzen enthalten können (Kompetenzstruktur- oder Kompetenzkomponentenmodelle) wie auch Informationen zu unterschiedlichen Ausprägungen oder Entwicklungsstufen (auch bezeichnet als Kompetenzniveaus) von Kompetenzen (Kompetenzstufen- oder Kompetenzentwicklungsmodelle) (Klieme et al. 2003; Schecker & Parchmann, 2006). Als Prinzipien für die Entwicklung von Kompetenzen können dabei, neben der Akkumulation von Wissen, eine zunehmende Prozeduralisierung des Wissens, wachsende Vernetzung des Wissens sowie der Aufbau von abstraktem und Meta-Wissen beschrieben werden (Klieme et al., 2003).

Weiterhin wird je nach zugrunde gelegter Fundierung zwischen normativen und deskriptiven Modellen unterschieden (Schecker & Parchmann, 2006). Normative Modelle beschreiben die notwendigen Voraussetzungen zur Lösung von Problemen in einem Anforderungsbereich und sollten eine solide theoretische Fundierung sowohl in fachlichen Bildungszielen wie auch

psychologischen Lerntheorien aufweisen, wohingegen deskriptive Modelle die typischen Ausprägungen dieser Voraussetzungen bei Lernenden beschreiben und empirisch abgesichert sein sollten (Schecker & Parchmann, 2006). Angestrebt werden sollte ein Zusammenwachsen normativer und deskriptiver Modelle, bei denen tatsächlich vorhandene Kompetenzen bei der Formulierung realistischer Teilziele als Orientierung genommen werden und normative Erwartungen als Zielformulierungen dienen können (Schecker & Parchmann, 2006).

Auf Basis der in den Modellen definierten Kompetenzen können dann Instrumente zur Erfassung entwickelt werden, wobei die angestrebte Funktion der erhobenen Informationen ein wichtiges Leitbild für die Umsetzung zur Messung darstellt. So können mit den Ergebnissen der Kompetenzerfassung individuelle Bildungsentscheidungen (Diagnostik), Entscheidungen über ganze Bildungssysteme (Assessment) oder Bewertungen von Maßnahmen hinsichtlich vordefinierter Kriterien (Evaluation) verknüpft sein (Klieme & Leutner, 2006). Der angemessene Auflösungsgrad der Instrumente wird von diesem Zweck mitbestimmt (Klieme & Leutner, 2006).

In den folgenden Abschnitten wird der aktuelle Forschungsstand zu Kompetenzmodellen in den Naturwissenschaften aufgearbeitet. Dabei wird vornehmlich auf den aktuellen Stand in Deutschland eingegangen, da die vorliegende Arbeit im Rahmen des deutschen Schulsystems verortet ist. Zudem wird, in Anbetracht der untersuchten Stichprobe dieser Arbeit, ein besonderes Augenmerk auf das Grundschulalter gerichtet.

## **2.1.2 Kompetenzmodelle in den Naturwissenschaften**

### **2.1.2.1 Normative Kompetenzmodelle in den Naturwissenschaften**

Für den Bereich der Naturwissenschaften wurden bereits zahlreiche normative Kompetenzmodelle beschrieben, die präskriptiv, ausgehend von fachlichen Bildungszielen und Befunden der lernpsychologischen Forschung, Anforderungen beschreiben, über die ein Lernender verfügen soll (vgl. z.B. American Association for the Advancement of Science, 1993; Gesellschaft für Didaktik des Sachunterrichts, 2002; National Science Education Standards, 1996; Qualifications and Curriculum Authority, 2000).

Beispielsweise hat Bybee (1997) in seiner Konzeption von naturwissenschaftlicher Grundbildung (*scientific literacy*, SL) zwischen vier Kompetenzstufen unterschieden: (1) nominale SL als Kenntnis naturwissenschaftlicher Begriffe, Fragen und Themen ohne tiefergehendes Verständnis, (2) funktionale SL als die Fähigkeit zur richtigen Verwendung naturwissenschaftlicher Ausdrücke, (3) konzeptuelle und prozedurale SL als ein Verständnis zentraler naturwissenschaftlicher Begriffe und Ideen und ihrer Beziehung zur gesamten Disziplin sowie ein Verständnis grundlegender Methoden und Verfahren und (4) multidimensionale SL als Verständnis des naturwissenschaftlichen Denkens und der Entwicklung naturwissenschaftlicher Ideen sowie ihrer gesellschaftlichen Bedeutung. Obwohl der Autor in diesem Modell noch nicht zwischen Inhaltsbereichen differenziert hat, ist er davon ausgegangen, dass die Entwicklung über die Stufen

domänenspezifisch geschieht und sich eine Person bzgl. zweier unterschiedlicher Domänen auf unterschiedlichen Entwicklungsstufen befinden kann (Bybee, 1997). Mit seinem Modell hat der Autor einen wichtigen Schritt in Richtung einer einheitlichen Definition von SL geleistet, die als eines der wichtigsten Ziele von Bildungssystemen national wie auch international schon lange diskutiert wird (Bybee, 1997).

Zur Normierung des Schulsystems in Deutschland wurden die Nationalen Bildungsstandards für die Naturwissenschaften in Bezug auf die Schulfächer Physik, Biologie und Chemie für den mittleren Schulabschluss, das Ende der zehnten Klasse, ausformuliert (Sekretariat der Ständigen Konferenz der Kultusminister, 2005). Dabei wurden für jeden der Inhaltsbereiche (Physik, Biologie, Chemie) vier Kompetenzbereiche unterschieden. Diese sind Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung. Der Bereich Fachwissen wird weiter in Basiskonzepte untergliedert, für das Fach Physik beispielsweise Materie, Wechselwirkung, System und Energie. Desweiteren werden drei kognitive Anforderungsbereiche für jeden Kompetenzbereich differenziert, für den Kompetenzbereich Fachwissen z.B. Wissen wiedergeben, Wissen anwenden und Wissen transferieren und verknüpfen. Zur Konkretisierung der Bildungsstandards wurden zu allen aufgespannten Bereichen Aufgabenbeispiele und ein Erwartungshorizont beschrieben (Sekretariat der Ständigen Konferenz der Kultusminister, 2005). Neben der inhaltlichen Dimension der Bildungsstandards, die durch die drei Naturwissenschaften abgedeckt ist, wird außerdem eine Handlungsdimension angenommen, die grundlegende Elemente der naturwissenschaftlichen Erkenntnisgewinnung umfasst. Die Überprüfung der Bildungsstandards und die Entwicklung von abgestimmten Messverfahren sind bereits angelaufen (vgl. Kauertz, Fischer, Mayer, Sumfleth & Walpuski, 2010; siehe auch Kapitel 2.1.2.3).

Für die Grundschule in Deutschland wurden Bildungsziele im sog. Perspektivrahmen Sachunterricht (Gesellschaft für Didaktik des Sachunterrichts, 2002) festgehalten. In diesem sind fünf Perspektiven (sozial- und kulturwissenschaftlich, raumbezogen, naturbezogen, technisch, historisch) aufgespannt worden, unter denen Themen und Inhalte für den Unterricht unter Berücksichtigung der Lebenswelt der Kinder und der Anschlussfähigkeit in weiterführenden Schulen ausgewählt werden können. Für die naturbezogene bzw. naturwissenschaftliche Perspektive werden inhalts- und verfahrensbezogene Kompetenzen für die ersten beiden und abschließenden beiden Klassenstufen unterschieden.

Seit der Formulierung des Perspektivrahmens ist dieser bei der Umsetzung von Lehrplänen in den Ländern der Bundesrepublik Deutschland berücksichtigt worden. So sind im Lehrplan des Landes Nordrhein-Westfalen für den Sachunterricht (Ministerium für Schule und Weiterbildung des Landes Nordrhein – Westfalen, 2008) zu unterrichtende Sachverhalte fünf Bereichen zugeordnet worden, die den Perspektiven des Perspektivrahmens entsprechen und denen jeweils unterschiedliche Schwerpunkte zugeordnet werden. Für alle Bereiche und Schwerpunkte wurde

ausgeführt, welche Kompetenzen die Schüler am Ende der Schuleingangsphase und am Ende der vierten Klasse erreicht haben sollen.

Da die hier beschriebenen normativen Modelle noch keiner empirischen Überprüfung unterzogen wurden, ist unklar, inwieweit ihre normativen Ansprüche von den betroffenen Lernenden erfüllt werden können und ob die vorgeschlagenen Dimensionen die tatsächliche Struktur der jeweiligen Kompetenzen wiedergeben.

#### *2.1.2.2 Deskriptive Kompetenzmodelle in den Naturwissenschaften*

Den normativen stehen deskriptive Kompetenzmodelle gegenüber, in denen, z.B. mit Hilfe der Methoden des *proficiency scaling*, aufgrund empirischer Daten die Stufen oder Struktur von Kompetenzen beschrieben werden. Diese Methode wurde vornehmlich bei internationalen Schulleistungstudien, wie z.B. PISA 2000 bis 2009 (Prenzel, Rost, Senkbeil, Häußler & Klopp, 2001; Klieme, Artelt & Hartig, 2010; Rönnebeck, Schöps, Prenzel & Hamann, 2008; Rost, Prenzel, Carstensen, Senkbeil & Groß, 2004) eingesetzt.

Die erste großangelegte Studie zur Beschreibung naturwissenschaftlicher Kompetenz in der Grundschule unter Beteiligung Deutschlands war die *Internationale Grundschul-Lese-Untersuchung* (IGLU; Bos, Lankes, Prenzel, Schwippert, Walther & Valtin, 2003). Im Rahmen dieser Untersuchung wurden die naturwissenschaftlichen Kompetenzen von Schülern am Ende der Grundschulzeit in einer nationalen Erweiterung untersucht (Lankes et al., 2003; Prenzel, Geiser, Langeheine & Lobemeier, 2003). Der Aufgabenentwicklung für diese Studie lag ein theoretisches Rahmenmodell zu Grunde, in welchem zwischen unterschiedlichen Stoffgebieten (Biologie, Physik und Chemie, Erde und Umwelt) und unterschiedlichen Kompetenzanforderungen (Wissensreproduktion, Konzeptanwendung, Verständnis, Prozesswissen) unterschieden wurde. Die empirisch erfassten Kompetenzen der Schüler wurden fünf Kompetenzstufen zugeordnet, die von einfacher Reproduktion naturwissenschaftlicher Bezeichnungen bis hin zu beginnendem naturwissenschaftlichen Denken und dem Anwenden einfacher naturwissenschaftlicher Konzepte reichten. Diese Kompetenzstufen wurden anhand der empirischen Schwierigkeitsverteilung der Aufgaben bei einer eindimensionalen Skalierung und ihren inhaltlichen Anforderungen festgelegt und beschrieben. Darüber hinaus erfolgte eine Charakterisierung jeder Stufe mit Aufgabenbeispielen.

Dieses Vorgehen entspricht auch dem des *Programme for International Student Assessment* (PISA; PISA-Konsortium Deutschland, 2007; 2008), in dem naturwissenschaftliche Kompetenzen von Sekundarschülern im Fokus des Interesses standen. Grundlage der Aufgabenentwicklung war hier eine Unterscheidung von deklarativem naturwissenschaftlichem Wissen und prozeduralem Wissen über Naturwissenschaften. Die Alltagstauglichkeit des Wissens sollte durch die flexible Anwendung in unterschiedlichen Kontexten geprüft werden (Rönnebeck et al., 2008). Die fünf unterschiedlichen Kontexte waren Gesundheit, natürliche Ressourcen, Umwelt, Gefahr und Risiken

sowie Grenzen von Naturwissenschaften und Technik. Eine Zuordnung zu nachträglich anhand der empirischen Aufgabenschwierigkeiten festgelegten Kompetenzstufen wurde auch in dieser Studie vorgenommen. Die beschriebenen Stufen reichen von einem eingeschränkten Wissen bis hin zu konsistenter Anwendung von naturwissenschaftlichem Wissen und Wissen über die Naturwissenschaften (Prenzel et al., 2001).

Bei der zweiten Runde der *Trends in International Mathematics and Science Study* (TIMSS; Bos, Bensen, Baumert, Prenzel, Selter & Walther, 2008) hat Deutschland ebenfalls an einer Studie zur Untersuchung der Kompetenzen von Grundschulern teilgenommen. Die Rahmenkonzeption für die Aufgabenentwicklung sah hier drei Inhaltsbereiche (Biologie, Physik, Geographie), die weiter in Themengebiete differenziert wurden, sowie drei kognitive Anforderungsbereiche (Reproduzieren, Anwenden, Problemlösen) vor (Bensen, Lintorf, Bos & Frey, 2008). Auf der empirischen Kompetenzskala wurden in dieser Studie fünf Kompetenzstufen unterschieden, die von rudimentärem Anfangswissen bis zu beginnendem naturwissenschaftlichem Denken reichten (Wittwer, Saß & Prenzel, 2008).

#### *2.1.2.3 Naturwissenschaftliche Kompetenzmodelle mit theoretischer Begründung und empirischer Absicherung*

Die nachträgliche Zuweisung von Kompetenzstufen, welche die Grundlage der Definition von Kompetenzstufen in den Modellen war, die im vorgehenden Kapitel 2.1.2.2 beschrieben wurden, ist allerdings als unzureichend kritisiert worden. Erstrebenswert ist die Zusammenführung eines theoretisch begründeten Vorgehens mit empirischer Absicherung, bei der eine normative Beschreibung von Kompetenz als Ausgangspunkt bei der Entwicklung von Aufgaben und bei der kritischen datengeleiteten Revision des Modells dienen kann (Schecker & Parchmann, 2006; Wilson, 2005). Für die Naturwissenschaften ist ein solches theoriegeleitetes empirisches Vorgehen in Teilen bereits geleistet worden.

Als Rahmenmodell für die Planung von empirischen Studien haben Schecker und Parchmann (2006) ein Modell mit den Dimensionen Inhaltsbereich/Basiskonzept, Handlung/Prozess, Kontext, Ausprägung und kognitive Anforderung vorgeschlagen. Die Autoren haben bei der Formulierung des Modells sowohl normative Modelle wie auch bereits vorliegende empirische Befunde über Schülerkompetenz berücksichtigt und auf eine Anschlussfähigkeit an die Nationalen Bildungsstandards geachtet (Schecker & Parchmann, 2006). Mit der Dimension Kontext ist die Abhängigkeit von Schülererklärungen vom jeweiligen Kontext, wie sie aus anderen Forschungsbereichen bekannt ist (vgl. Kap. 2.2.3), berücksichtigt worden.

Kauertz (2008) hat in seiner Arbeit zur Strukturierung der Kompetenz im Bereich Physik die Dimensionen kognitive Aktivität, Komplexität und Leitidee beschrieben. Alle drei Dimensionen beziehen sich auf Merkmale der Inhaltsstruktur von Aufgaben (Kauertz, 2008): Die kognitive Aktivität gibt an, ob in einer Aufgabe das Erinnern, Strukturieren oder Explorieren gefordert ist.

Entsprechend der Bildungsstandards für das Fach Physik umfasst die Dimension Leitidee die Basiskonzepte Energie, Wechselwirkung, System und Materie (siehe Kapitel 2.1.2.1). Die Komplexität zeigt an, welche Struktur des Inhalts in der Lösung einer Aufgabe erwartet wird, wobei sechs hierarchische Kompetenzniveaus unterschieden worden sind: (1) ein Fakt, (2) mehrere Fakten, (3) ein Zusammenhang, (4) mehrere unverbundene Zusammenhänge, (5) mehrere verbundene Zusammenhänge und (6) ein übergeordnetes Konzept (Kauertz, 2008). Der Autor konnte einen Einfluss der Komplexität auf die empirische Schwierigkeit entsprechend entwickelter Aufgaben nachweisen und diesen Effekt auch gegenüber anderen, formalen Eigenschaften der Aufgaben absichern (Kauertz, 2008).

Eine Weiterentwicklung des Modells von Kauertz (2008) haben Viering, Fischer und Neumann (2010) beschrieben. Statt der Lösungskomplexität von Aufgaben haben sie die Aufgabenkomplexität modelliert, die sich aus der Differenz der Komplexität der im Aufgabentext zur Verfügung gestellten Information und der Komplexität der erwarteten Lösung ergibt. Zusätzlich zu den von Kauertz (2008) vorgeschlagenen Dimensionen haben Viering und Kollegen (2010) mit der Konzeptentwicklung neben hierarchischen Kompetenzstufen außerdem eine Dimension benannt, die sich explizit auf die Entwicklung von Kompetenzen bei Individuen bezieht. Sie konnten in einer Pilotstudie ein hypothesenkonformes Ansteigen der Aufgabenschwierigkeiten mit den a priori angenommenen Stufen der Kompetenzentwicklung zeigen (Viering et al., 2010).

Auch für die Evaluation der Nationalen Bildungsstandards (NBS) in den Naturwissenschaften (siehe 2.1.2.1) wurde eine Beschreibung von Kompetenzstufen in Anlehnung an das Modell von Kauertz (2008) gewählt. Die inhaltliche Strukturierung der NBS mittels Kompetenzbereichen und Basiskonzepten wurde für die Operationalisierung der untersuchten Kompetenzen übernommen und weiter ausdifferenziert. Als zweite Dimension sind kognitive Prozesse (reproduzieren, selektieren, organisieren, integrieren) beschrieben worden (Walpuski et al., 2010). Zur Beschreibung von Kompetenzniveaus wurde die Komplexität der Aufgaben gewählt. Im Vergleich zum Modell von Kauertz (2008) sind allerdings nur fünf statt der ursprünglichen sechs Stufen der Komplexität angenommen worden (Kauertz et al., 2010). Die Durchführung der Evaluation des Modells war für das Jahr 2012 geplant.

Der Bereich der Grundschule ist in dem Schweizer Projekt *Harmonisierung obligatorische Schule* (HarmoS; Labudde, 2007) bearbeitet worden. Ziel von HarmoS war die Entwicklung einheitlicher Standards für das Ende des 2., 6. und 9. Schuljahres im Schweizer Schulsystem (Labudde, 2007). Für die Naturwissenschaften ist in dem Kompetenzmodell zwischen zehn Handlungsaspekten (z.B. Ordnen, strukturieren und modellieren, Fragen und untersuchen), neun Themenbereichen (z.B. Planet Erde, Bewegung, Kraft, Energie) und unterschiedlichen Niveaus unterschieden worden (Zeyer, Adamina, Gingins & Labudde, 2008). Umfangreiche empirische Untersuchungen mit Aufgaben, die auf Grundlage dieses Modells entwickelt wurden, lieferten die Grundlage für die

Formulierung von Basisstandards in allen untersuchten Aspekten (Labudde & Adamina, 2008). Dabei zeigten sich allerdings hohe Zusammenhänge zwischen den Handlungsaspekten bzw. Themenbereichen, welche die Ausdifferenzierung der Kompetenz in so viele Aspekte fragwürdig erscheinen lassen (Labudde & Adamina, 2008).

Für die Grundschule wurden weitere Kompetenzmodelle für einzelne Teilbereiche der Naturwissenschaften formuliert und empirisch überprüft. So hat Grygier (2008) ein Modell für das Wissenschaftsverständnis vorgeschlagen, in welchem drei Levels der Kompetenz, sowie jeweils Zwischenlevel beschrieben worden sind. Darüber hinaus hat Hammann (2004) ein Modell zur Experimentierkompetenz ausgearbeitet. Sommer (2006) hat ein Modell zur Systemkompetenz in der Biologie vorgelegt. Außerdem sind das Modell von Schmidt und Schecker (2006) zum Inhaltsbereich Energie sowie die Arbeit zur Erfassung der Teilkompetenz Bewerten, Entscheiden und Reflektieren von Bögeholz und Kollegen (Bögeholz, Höble, Langelt, Sander & Schlüter, 2004) zu nennen.

#### 2.1.2.4 Das Projekt Science-P

Ebenfalls für die Grundschule wird im Projekt *Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule (Science-P)*<sup>1</sup> ein zweidimensionales Struktur-Niveau-Modell untersucht (Hardy et al., 2010). Da die vorliegende Arbeit in diesem Projekt verortet ist, wird es in diesem Kapitel ausführlich vorgestellt.

Im Projekt Science-P wurden in Anlehnung an die Befunde der Conceptual Change-Forschung, nach denen Schüler bereits vor dem Unterricht eigene Vorstellungen über naturwissenschaftliche Phänomene halten, die während des Lernprozess nach und nach in angemessenere Erklärungen überführt werden (vgl. 2.2.1), allgemeine Kompetenzniveaus formuliert. Dabei wurde von drei hierarchischen Niveaus der Kompetenz ausgegangen (Hardy et al., 2010): Das unterste Niveau bilden *naive Vorstellungen*, die häufig nicht mit wissenschaftlichen Erklärungen vereinbar sind. Auf dem mittleren Niveau kann es während des graduellen Prozesses der Umstrukturierung der *naiven Vorstellungen* zu sog. *Zwischenvorstellungen* kommen, die in einer größeren Auswahl von Situationen einer Überprüfung standhalten und bereits richtige Aspekte enthalten können. Das oberste Niveau bilden die *wissenschaftlichen Vorstellungen*, welche die wissenschaftlich adäquate Sichtweise in altersangemessener Sprache darstellen.

Auf der inhaltlichen Ebene wurde in diesem Projekt in Anlehnung an Bybees (1997) Konzeption von prozeduraler und konzeptueller *scientific literacy* (siehe 2.1.2.1) zwischen dem *Wissen über*

---

<sup>1</sup> Das Projekt Science-P wurde ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293) und gemeinsam von den Arbeitsgruppen Sodian/Mayer (München), Hardy/Jurecka/Kempert (Frankfurt), Koerber/Osterhaus (Freiburg), Schwippert (Hamburg) und Möller/Pollmeier/Tröbst (Münster) durchgeführt.



*Naturwissenschaften* mit den Komponenten Wissenschaftsverständnis und Methodenkompetenz und dem *Naturwissenschaftlichen Wissen* mit den exemplarischen Komponenten Schwimmen/Sinken und Verdunstung/Kondensation unterschieden. Diese Themenbereiche wurden wiederum weiter in Komponenten bzw. Konzepte ausdifferenziert. So wurden für den Themenbereich Schwimmen/Sinken die Komponenten Dichte, Verdrängung und Auftrieb und für den anderen Themenbereich die Komponenten Verdunstung und Kondensation untersucht. Dieses Kompetenzmodell aus Kompetenzniveaus und inhaltlichen Bereichen war Grundlage der Aufgabenentwicklung (vgl. Kap. 4.3.1.1).

Es konnte zur Erfassung der Kompetenzen ein valides Messinstrument entwickelt werden (Hardy et al., 2010; Kleickmann et al., 2010; Pollmeier, Hardy, Koerber & Möller, 2011). Erste empirische Ergebnisse zur Struktur der untersuchten Kompetenz deuten darauf hin, dass sie inhaltspezifisch stark variieren (Kleickmann, Hardy, Pollmeier & Möller, 2011).

#### *2.1.2.5 Zusammenfassung und Fazit*

Die in Kapitel 2.1.2.1 beschriebenen normativen Kompetenzmodelle haben Anforderungen formuliert, die Lernende erreichen sollen. Während der Begriff der *scientific literacy* von Bybee (1997) einen allgemeinen Orientierungsrahmen bietet, der für viele Diskussionen um Bildungsziele zentral war und ist, wurden mit den Nationalen Bildungsstandards (Sekretariat der Ständigen Konferenz der Kultusminister, 2005), dem Perspektivrahmen Sachunterricht (Gesellschaft für Didaktik des Sachunterrichts, 2002) und dem Lehrplan des Landes Nordrhein-Westfalen (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008) exemplarisch drei Kompetenzmodelle vorgestellt, die hohe Relevanz für das aktuelle unterrichtliche Geschehen in Deutschland haben, ohne zuvor einer umfassenden empirischen Überprüfung unterzogen worden zu sein. Insgesamt sind alle genannten Modelle sehr allgemein formuliert und sind in ihrer jeweiligen Ausschärfung noch nicht direkt einer Messung zugänglich. Zwar wurden auch teilweise bereits Aufgabenbeispiele für die Erfassung von Kompetenz bzw. für die Konkretisierung der gefragten Kompetenzen beschrieben; für eine ausführliche Validierung der Modelle steht jedoch noch viel Arbeit hinsichtlich der Operationalisierung und Umsetzung in Testaufgaben an.

Demgegenüber sind die in Kapitel 2.1.2.2 beschriebenen Kompetenzstufenmodelle alle auf Grundlage einer umfassenden empirischen Datenbasis entstanden. In den beschriebenen Studien wurden ausführliche Definitionen der untersuchten naturwissenschaftlichen Kompetenz für die Aufgabenentwicklung vorgenommen und die erhobenen Daten wurden psychometrischen Skalierungen unterzogen. Auf den so entstandenen Skalen wurden Kompetenzstufen definiert, die sich an der Anordnung von Aufgaben unterschiedlichen Inhalts entlang der jeweiligen Skala orientierten. Den so entstandenen Post-hoc-Beschreibungen unterschiedlicher Ausprägungen der Kompetenz fehlt in der Regel jedoch eine theoretische Fundierung. Die eindimensionalen Skalen, die ein zentrales Resultat der Studien bilden, erlauben einen unproblematischen Vergleich der

durchschnittlichen Fähigkeit der Schüler verschiedener Nationen. Allerdings ist ein direkter Schluss auf individuelle Leistungsprofile, z.B. auf relative Defizite in einem Fachgebiet, nicht möglich. Dadurch ist eine direkte Nutzung der Ergebnisse für den Unterricht, um beispielsweise Schwächen gezielt entgegenzusteuern, schwierig.

Die empirische Validierung theoretisch fundierter Kompetenzmodelle wurde in den in den Kapiteln 2.1.2.3 und 2.1.2.4 vorgestellten Arbeiten angestrebt. Schecker und Parchmann (2006) haben ein umfassendes Modell mit Konkordanz zu den NBS vorgeschlagen. In den Modellen von Kauertz (2008), von Viering und Kollegen (Viering et al., 2010) und zur Evaluation der Bildungsstandards (Kauertz et al., 2010; Walpuski et al., 2010) wurde jeweils eine Dimension der Komplexität zur Graduierung der Kompetenz bzw. der Aufgabenschwierigkeit beschrieben. In den Projekten HarmoS (z.B. Labudde & Adamina, 2008) und Science-P (z.B. Hardy et al., 2010), welche den Grundschulbereich betreffen, wurden hingegen Kompetenzniveaus beschrieben. In Science-P wurden diese explizit aus der Conceptual Change-Forschung abgeleitet (Hardy et al., 2010; Kleickmann et al., 2010).

Anhand der beschriebenen Modelle wird deutlich, dass die Kompetenzmodellierung in den Naturwissenschaften in Deutschland erst angelaufen ist. Die notwendigen Studien inklusive der notwendigen theoretischen Vorarbeiten sind sehr zeitintensiv, und zur Umsetzung der geeigneten statistischen Modelle, wie z.B. Modellen der *Item Response Theory* (Klieme & Leutner, 2006), sind hohe Stichprobenumfänge unabdingbar. Zudem ist häufig ein iteratives Vorgehen erforderlich, wenn das postulierte Kompetenzmodell anhand der Daten aus Vorstudien nicht bzw. nicht vollständig bestätigt werden kann. Darüber hinaus stellt es sich als schwierig heraus, Modelle für die Naturwissenschaften zu formulieren, die diesen breiten Bereich angemessen abdecken, ohne die Relevanz für den Unterricht durch eine zu starke Verallgemeinerung zu verlieren oder durch eine zu starke Zergliederung in thematische Unterbereiche mit zu feiner Auflösung vorzugehen. Damit Kompetenzmodellierung und -diagnostik auch für gezielte Veränderungen des unterrichtlichen Geschehens nutzbar gemacht werden können, sollten die Ergebnisse nicht nur für eine summative Nutzung der Daten, d.h. zur abschließenden Beurteilung der Lernzielerreichung, sondern auch zur formativen Messung, d.h. zur Bestimmung individueller Lernstände im Unterrichtsverlauf, geeignet sein (Black & Wiliam, 2009; Maier, 2010). Für die Fachdidaktiken sind in erster Linie Modelle der Entwicklung von Kompetenz relevant (Beinbrech & Möller, 2008). Dies ist für den Bereich der Grundschule in Deutschland erst in Ansätzen geleistet.

Die vorliegende Arbeit befasst sich mit der Modellierung naturwissenschaftlicher Kompetenz in der Grundschule. Das vorhandene Wissen bei Grundschulern zu diesem Bereich sowie die Veränderung dieses Wissens während des Lernens wurden intensiv unter dem Paradigma der Forschung zu Conceptual Change und Schülervorstellungen erforscht (vgl. Kap. 2.2.1). Die zahlreichen Erkenntnisse, die aus dieser Forschungsrichtung vorliegen, können für die Formulierung, Umsetzung und Überprüfung eines Modells der Entwicklung

naturwissenschaftlicher Kompetenz in der Grundschule genutzt werden. Sie können potentiell Hinweise auf die Struktur des Wissens liefern und mögliche Einflussfaktoren benennen. Die Validität der Erfassung naturwissenschaftlicher Kompetenz in der Grundschule könnte also von Befunden der Conceptual Change-Forschung profitieren. Im folgenden Kapitel 2.2 wird daher der Bereich Schülervorstellungen und Conceptual Change in den Naturwissenschaften aufgearbeitet. Ein besonderer Fokus wird dabei auf Faktoren gelegt, die eine Validierung von Aufgaben zur Erfassung naturwissenschaftlicher Kompetenz in der Grundschule unterstützen können. Dazu gehört insbesondere der Einfluss des Kontexts auf Schülervorstellungen. Wie bereits in Kapitel 2.1.1 angedeutet, ist der Kontext ein entscheidendes Element bei der Operationalisierung von Kompetenzen. Auch im Rahmen der Conceptual Change-Forschung wurde der Kontext als kritisches Element identifiziert. Diese Übereinstimmung kann zur Verbesserung der Validität bei der Kompetenzmodellierung genutzt werden.

## **2.2 Schülervorstellungen in den Naturwissenschaften**

### **2.2.1 Schülervorstellungen und Conceptual Change**

Grundschul Kinder entwickeln oft auch ohne systematischen Unterricht durch ihre Alltagserfahrungen vielfältige Vorstellungen zu naturwissenschaftlichen Themen (Driver, 1989). Seit den 1970er Jahren hat sich Forschung in der Lernpsychologie und der naturwissenschaftlichen Bildung mit der Untersuchung dieser Vorstellungen beschäftigt. Häufig verwendete Begriffe für dieses Phänomen sind im deutschen Sprachraum *Schülervorstellungen*, *Präkonzepte*, *naive Theorien* oder *Vorerfahrungen* (Möller, 2010), im englischen Sprachraum werden die vorunterrichtlichen Schülerkonzepte häufig als *preconceptions*, *misconceptions*, *intuitive* oder *naive theories*, *naive conceptions* oder *alternative frameworks* bezeichnet (diSessa, 2006; Driver & Easley, 1978; Novak, 1987; Tytler, 1998). Im Gegensatz zu den Vorstellungen vor dem Unterricht spricht Möller (2010) bei bereits durch Instruktion beeinflussten Konzepten von *Postkonzepten*. Für die Grundschule wurde der Begriff von Carey (1985) geprägt.

Trotz der Vielfalt der verwendeten Bezeichnungen, finden sich Übereinstimmungen in den Charakterisierungen der vorunterrichtlichen Schülervorstellungen: Es werden darunter gedankliche Entwürfe verstanden (Möller, 2010), die jeweils spezifisch für eine inhaltliche Domäne sind (Driver, 1989). Diese sind häufig nicht mit wissenschaftlichen Vorstellungen vereinbar und nur schwer durch traditionelle Formen des Unterrichts zu verändern (Treagust & Duit, 2008). Die Vorstellungen von Schülern treten in großer Vielfältigkeit auf, wobei aber Regularitäten bzw. Gemeinsamkeiten über verschiedene kulturelle Kontexte und Altersgruppen hinweg festgestellt werden konnten (Scott, Asoko & Leach, 2007). Die Wurzeln der Schülerkonzepte können in alltäglichen, außerschulischen Erfahrungen der Schüler mit ihrer physikalischen Umwelt (Driver & Easley, 1978; Engel Clough & Driver, 1985) oder in Formulierungen des alltäglichen

Sprachgebrauchs liegen (Garnett, Garnett & Hackling, 1995). Die gemachten Erfahrungen bzw. alltagssprachlichen Formulierungen werden von den Schülern interpretiert und die Deutungen für Erklärungen von Phänomenen bzw. Gegebenheiten genutzt (Driver & Easley, 1978). Beispielsweise führt die tägliche Erfahrung von „oben und unten“ dazu, dass v.a. junge Kinder ein mentales Modell der Erde haben, welches von ihnen in Einklang mit eben dieser Erfahrung gebracht werden kann (z.B. Vosniadou & Brewer, 1992). Gerade die Assoziation bestimmter Vorstellungen mit persönlichen Erfahrungen kann die o.g. Robustheit gegen Veränderung durch Unterricht dieser verstärken (Dole & Sinatra, 1998).

Das Lernen neuer, angemessenerer Vorstellungen bedarf einer Veränderung der vorhandenen. Dieser Prozess ist – gerade wegen der Verwurzelung in eigenen Erfahrungen - langwierig und graduell und Schüler können dabei individuell unterschiedliche Lernwege beschreiten, die sie über teilrichtige Ideen zu angemesseneren Vorstellungen führen (z.B. Vosniadou, Ioannides, Dimitrakopoulou & Papademetriou, 2001). Viele Autoren unterscheiden dabei zwischen unterschiedlichen Arten der Konzeptveränderung. Hierbei reichen die beschriebenen Mechanismen auf einem angenommenen Kontinuum von einer Anreicherung des bestehenden Wissen über eine Ausdifferenzierung des Konzeptgefüges und das Ausbilden neuer Verknüpfungen zwischen Konzepten oder einer Integration von bestehendem und zu lernendem Wissen bis hin zur Ablösung ganzer Teile des konzeptuellen Gefüges (vgl. z.B. Dole & Sinatra, 1998; Vosniadou & Ioannides, 1998). Als tatsächlicher Konzeptwechsel (englisch: *conceptual change*) und zentraler Lernmechanismus wird dabei von den meisten Autoren nur letzteres angesehen (Dole & Sinatra, 1998).

Für die naturwissenschaftliche Bildung und Erziehung wurde das Conceptual Change-Modell von Posner, Strike, Hewson und Gertzog (1982) geprägt. Sie griffen für die Entwicklung des Modells auf Theorien der kognitiven Psychologie zurück, die sich schon länger mit dem Lernen bzw. der aktiven Konstruktion von Wissen beschäftigten und zogen Parallelen zur Geschichte und Philosophie der Naturwissenschaften (Dole & Sinatra, 1998; Posner et al., 1982). In ihrer einflussreichen Arbeit unterschieden sie zwischen der Assimilation, bei der unbekannte Phänomene mit den bereits existierenden Konzepten erklärt werden, und der Akkommodation, der Reorganisation oder dem Austausch der bestehenden Konzepte (Posner et al., 1982). Beiden Prozessen ist gemein, dass Wissen nicht einfach vom Lerner auf den Schüler transferiert werden kann, sondern aktiv konstruiert werden muss, wobei die Akkommodation als der zentrale Konzeptwechsel-Mechanismus angesehen wurde (Posner et al., 1982). Damit eine solche Akkommodation auftritt, müssen nach Posner und Kollegen (1982) folgende vier Bedingungen erfüllt sein: Der Lerner muss mit seinen vorhandenen Konzepten unzufrieden sein. Das zu lernende, neue Konzept sollte hingegen verständlich, plausibel und fruchtbar sein. Ähnlich wie Posner und Kollegen (1982) hat auch Vosniadou (1994) zwischen der Anreicherung der bestehenden Wissensstrukturen, falls das zu lernende Wissen konsistent mit dem bereits

verfügbaren ist, was nur zu einer geringfügigen Änderung der Vorstellungen führt (*weak revision*), und der Revision, falls das zu lernende Wissen inkonsistent mit dem bereits gelernten ist, was eine drastische Veränderung darstellt (*radical restructuring*), unterschieden.

Das Vorwissen spielt für den Lernprozess in allen Fällen eine entscheidende Rolle, da es den Rahmen für die Wahrnehmung, Verarbeitung und Interpretation von neuen Informationen bildet (Duit, 1999; Chinn & Brewer, 1993; Ausubel, 1968 nach Scott, Asoko & Leach, 2007) und es so zu einer Interaktion zwischen bestehendem und neuem Wissen kommen kann (Tytler, 1998). Dabei kann das Vorwissen sowohl hinderlich für das zu erwerbende Wissen sein als auch als produktiver Anknüpfungspunkt dienen. Dem Lerner wird bei diesem Prozess der Konzeptveränderung eine aktive Rolle bei der Konstruktion des Wissens zugeschrieben (Treagust & Duit, 2008). Häufig ist dabei festzustellen, dass Schüler neue Konzepte schon annehmen und diese parallel zu ihren vorschulischen Konzepten halten (Duit, 1999) oder sogar durch die Integration von naiven und erlernten Konzepten neue, synthetische Modelle bilden (Vosniadou & Ioannides, 1998).

Neben der rein kognitiven Betrachtung haben andere Autoren den Einfluss weiterer Größen auf den Prozess der Konzeptveränderung betont: So können motivationale, affektive und Kontextbedingungen (Pintrich, Marx & Boyle, 1993), aber auch soziale Faktoren (Cole, John-Steiner, Scribner & Sauberman, 1978) das Lernen bzw. den Prozess der Konzeptveränderung mitbestimmen. Darüber hinaus werden metakognitive Fähigkeiten als positive Voraussetzung für den Erwerb konzeptuellen Wissens erachtet (Caravita, 2001).

Die Fachdidaktik der Naturwissenschaften hat den Konzeptwechsel als Paradigma empirischer Lehr-Lernforschung genutzt und die Conceptual Change-Forschung hat hierbei wichtige Informationen über vorherrschende (Fehl-)Konzepte sowie günstige Lernbedingungen geliefert (Möller, 1999). In beiden Forschungsrichtungen herrscht Einigkeit darüber, dass das Vorwissen zu einem Themenbereich das weitere Lernen beeinflusst und dass der Lernende aktiv an der Umstrukturierung seines Wissens beteiligt sein muss (Duit, 1999; Möller, 2004). Als wichtig wird daher angesehen, dass die Lehrkraft die Schülervorstellungen für die Planung und Gestaltung des weiteren Unterrichts berücksichtigt (Möller, 2004).

Entsprechende konzeptwechselförderliche Prinzipien wurden aus Theorien und Befunden der Conceptual Change-Forschung abgeleitet und in Unterrichtsentwürfen umgesetzt (z.B. diSessa, 2008; Hardy, Jonen, Möller & Stern, 2006; Vosniadou et al., 2001). Der Erfolg solcher Lehr-Lern-Umgebungen zeigt die Relevanz, die diese Forschungsstränge für die Praxis haben. Eine theoriebasierte Modellierung und Beschreibung naturwissenschaftlichen Wissens, zur Definition von Lernzielen z.B. in Kompetenzmodellen (siehe 2.1.2), sollte daher die Befunde und Annahmen der Konzeptwechselforschung zur Struktur und Entwicklung von Schülervorstellungen berücksichtigen.

Allerdings werden je nach theoretischer Perspektive auch unterschiedliche Empfehlungen für die Gestaltung von Lernumgebungen gemacht, die von einer Sequenzierung des Unterrichtens

entsprechend relevanter vorläufiger Modelle (Vosniadou et al., 2001) bis hin zu dem Herausstellen wichtiger Verbindungen unterschiedlicher Wissens Elemente (diSessa, 2008) reichen.

Desweiteren sind trotz der bislang intensiven Bearbeitung von Schülervorstellungen und Konzeptveränderungen einige Fragen auf diesem Gebiet bislang noch ungeklärt: V.a. über die Organisation und Struktur der vorunterrichtlichen Vorstellungen von Schülern bestehen unterschiedliche theoretische Annahmen, die im Folgenden dargestellt werden.

### **2.2.2 Kohärenz vs. Fragmentierung?**

Hinsichtlich der Beschaffenheit der mentalen Organisation von Schülervorstellungen werden unterschiedliche Ansätze vertreten, die sich v.a. hinsichtlich der relativen Kohärenz bzw. Fragmentierung, die dem initialen Wissen zugeschrieben wird, unterscheiden: Während die Vertreter der Kohärenz davon ausgehen, dass die vorunterrichtlichen Vorstellungen schon früh die Form in sich konsistenter, naiver Theorien annehmen, postuliert der Fragmentierungs-Ansatz, dass es sich dabei eher um eine lose organisierte Ansammlung einzelner Wissens Elemente handelt. In diesem Abschnitt werden die zentralen Theorien der jeweiligen Ansätze beschrieben und vergleichend gegenübergestellt.

Vosniadou und Kollegen (Vosniadou, 1994; Vosniadou et al., 2008) gehen von sich früh entwickelnden Rahmentheorien (*framework theories*) aus, die ein in sich kohärentes System von epistemologischen und ontologischen Vorannahmen darstellen. Auf deren Grundlage werden neue Informationen und Überzeugungen (*beliefs*), die durch Beobachtung und Interpretation der physikalischen Welt erworben werden, in sog. spezifischen Theorien (*specific theories*) organisiert. Die Erklärung und Vorhersage problemhaltiger Situation geschieht anhand von mentalen Modellen (*mental models*), die als analoge Repräsentationen der Theorien in den Situationen spontan generiert werden. Durch die Assoziation dieser mentalen Modelle mit den tiefverwurzelten Rahmentheorien und ihre wiederholte Konsultation in unterschiedlichen Situationen wird nach Vosniadou (1994) eine interne Kohärenz der Vorstellungen über Situationen hinweg erkennbar. Neu erlernte Erklärungen werden zunächst im spezifischen Lernkontext angewandt, bevor sie auf eine breitere Auswahl von Situationen generalisiert werden können (Ioannides & Vosniadou, 2002). Lernschwierigkeiten ergeben sich nach diesem Ansatz, wenn die Prinzipien der zu lernenden wissenschaftlichen Theorien nicht vereinbar mit den Prinzipien der naiven Rahmentheorien sind und so eine Veränderung tief verwurzelter Vorannahmen notwendig wird (Vosniadou et al., 2001). In diesem Fall wird durch eine Integration von naiven Vorstellungen mit neu zu lernenden wissenschaftlichen Vorstellungen in sog. synthetischen Modellen ein paralleles Halten naiver und wissenschaftlicher Erklärungen möglich (Vosniadou et al., 2008), was als Fragmentierung in Bezug auf die Angemessenheit der Vorstellungen interpretiert werden kann (Vosniadou, 2012). Diese ist allerdings nicht den Schülervorstellungen inhärent, sondern tritt erst

nach der Konfrontation mit inkompatiblen Konzepten auf, wenn z.B. Unterricht nicht optimal auf das vorhandene Vorwissen der Schüler abgestimmt ist (Vosniadou, 2012).

Chi, Slotta und de Leeuw (1994) gehen in ihrer Theorie davon aus, dass alles Wissen in ontologischen Kategorien organisiert wird, die Eigenschaften von ihnen zugeordneten Subkategorien und Konzepten beschreiben. Dabei sind die ontologischen Kategorien sowohl im Hinblick auf ihre Mitglieder wie auch die Attribute zueinander disjunkt. Die Subkategorien innerhalb einer ontologischen Kategorie teilen hingegen Eigenschaften (Chi et al., 1994). Entsprechend dieser hierarchischen Wissensorganisation geht Chi (2008) in ihrem Ansatz von drei möglichen Arten der Konzeptveränderung aus, die auf unterschiedlichen hierarchischen Niveaus angesiedelt sind (*individual beliefs, mental models, ontological categorization*). Im einfachsten Fall handelt es sich bei fehlerhaften Schülervorstellungen um einzelne Ideen, die durch einfaches Widerlegen korrigiert werden können. Die Transformation eines mentalen Modells im Sinne einer Sammlung in sich kohärenter, aber (teilweise) fehlerhafter Ideen kann ebenfalls durch das Widerlegen der kritischen (Kern-)Ideen gelingen. Wurde ein Konzept allerdings einer falschen ontologischen Kategorie zugewiesen, ist ein sog. *categorical shift* notwendig, der voraussetzt, dass der Lernende sich der fehlerhaften Kategorisierung bewusst ist und eine angemessenere Kategorie verfügbar ist. Besondere Lernschwierigkeiten treten dann auf, wenn das zu lernende Konzept inkompatibel mit der fehlerhaften Kategorie des Vorwissens ist. Diese Inkompatibilität führt laut Chi und Kollegen (1994) zu robusten und konsistenten naiven Konzepten.

Diese beiden Ansätze nehmen ein kohärentes, theorieähnlich organisiertes Wissen an, das in unterschiedlichen Situationen zu ähnlichen Vorhersagen und Erklärungen und somit zu einer Konsistenz der naiven Vorstellungen führen kann, obwohl auch Inkohärenz möglich ist. Den Ansätzen ist zudem gemein, dass einzelne Konzepte erst durch die Zuordnung zu einer Rahmentheorie bzw. ontologischen Kategorie einen Informationsgehalt erfahren (Vosniadou, 1994).

Demgegenüber stehen Ansätze, in denen die naiven Ideen als vielfältig, nicht theorieähnlich, sondern als unabhängige Fakten beschrieben werden (diSessa, Gillespie & Esterly, 2008). Beispielsweise beschreibt diSessa (1993; 2008) das anfängliche Wissen als eine unstrukturierte, nur lose organisierte Ansammlung einzelner Wissens Elemente (*phenomenological primitives, p-prims*). Die einzelnen Elemente erfahren in problemhaltigen Situationen eine kontextspezifische Aktivierung und können durch gemeinsame Aktivierung mehrerer Fragmente für theorieähnliche Erklärungen genutzt werden (diSessa, 2006). Die Wissens Elemente selbst können dabei Aktivierung in einer relativ großen Anzahl unterschiedlicher Situationen erfahren. Allerdings kann die Auswahl von Phänomenen, die zu einer ähnlichen Aktivierung führen, sich stark von der wissenschaftlichen Klassifikation derselben Phänomene hinsichtlich der relevanten (beispielsweise physikalischen) Konzepte unterscheiden (diSessa et al., 2004). Die Herausforderung beim Aufbau von wissenschaftlich angemessenem Wissen liegt darin, die angemessene Aktivierung der *p-prims*

und damit die Anwendung von adäquatem Wissen in unterschiedlichen Kontexten zu erreichen. Im Unterschied zu den oben beschriebenen Ansätzen wird in diesem Ansatz nicht davon ausgegangen, dass die *p-prims* einem größeren theoretischen System zugeordnet werden. Erst durch systematischen Unterricht werden neue Verbindungen zwischen den Fragmenten gebildet und eine Organisation in größeren, theorieähnlichen Strukturen möglich. Dies geschieht in hierarchischen Systemen, die zu einer stärkeren Systematisierung der Wissens und der Wahrnehmung führen (z.B. *coordination classes*: diSessa, 2006; diSessa & Sherin, 1998). Dennoch sind die einzelnen *p-prims* weder als zufällig noch als fragil anzusehen und können somit die Robustheit und auch die nach außen erscheinende Kohärenz von Vorwissen erklären (diSessa, 2008).

Sowohl in Ansätzen der Kohärenz wie auch in denen der Fragmentierung sind Wissens Elemente mit feiner Auflösung, d.h. Elemente mit hohem Zerlegungsgrad und einem geringen Umfang an enthaltener Information, zu finden: die *p-prims* im Ansatz von diSessa (2008), die *individual beliefs* bei Chi (2008) oder die Überzeugungen bei Vosniadou und Ioannides (1998), die in den spezifischen Theorien organisiert sind. Der Unterschied der Ansätze besteht in dieser Hinsicht darin, dass die Vertreter der Kohärenz davon ausgehen, dass diese einzelnen Elemente ihren Informationsgehalt erst durch die Zuordnung zu einer größeren Kategorie bzw. Theorie, erhalten, während die Fragmentierungs-Vertreter postulieren, dass die einzelnen Wissens Elemente selbsterklärend und direkt der Wahrnehmung zugänglich sind (diSessa, 2008; Vosniadou & Ioannides, 1998). Zudem sind in der Theorie der Fragmentierung die Wissens Elemente die zentralen Konstituenten des anfänglichen Wissens und stehen im Fokus der Betrachtung. Sie sind von großer Zahl und feiner Auflösung (diSessa, 2008). Die Vertreter der Rahmentheorie beziehen sich hingegen hauptsächlich auf eine geringere Anzahl umfassenderer naiver Theorien (Vosniadou & Ioannides, 1998). Von der Fragmentierungs-Seite dieser unterschiedlichen Theorien wird darüber hinaus die Kontextualität der Wissens Elemente expliziert und eine stärkere Berücksichtigung der Kontexte in den Untersuchungen zu Schülervorstellungen gefordert (diSessa, 2006).

Obwohl die Veränderung von Konzepten in allen Ansätzen als grundlegender Mechanismus des Lernens angesehen wird, ist dennoch jeweils unterschiedlich, welche Veränderungen als relevant erachtet werden und welche Folgerungen für die Gestaltung von Unterricht daraus gezogen werden. Für Vertreter des Kohärenzansatzes ist es entscheidend, eine Umstrukturierung des Wissens über eine gezielte Sequenzierung zu erlernender Theorien zu gestalten, wobei ein zentraler Weg dabei das Aufzeigen von Gegenevidenz zu den anfänglichen Theorien, verbunden mit dem Nahebringen fortgeschrittener Konzepte, ist (diSessa, 2006; Vosniadou et al., 2008). Hingegen betonen die Vertreter des Fragmentierungsansatzes die Bedeutung des Verdeutlichens von Beziehungen zwischen zunächst unverbundenen Wissens Einheiten sowie teilrichtiger Ansätze in diesen (diSessa, 2008) und sehen die bereits existierenden Wissens Elemente als produktive Bauteile des zu erwerbenden Wissens an (diSessa, 2006).



Welche Beschreibung der Organisationsform und Struktur des anfänglichen naturwissenschaftlichen Wissens angemessener ist, ist noch eine offene Frage. Die empirische Abgrenzung zwischen Kohärenz- und Fragmentierungs-Ansatz ist trotz der aufgezeigten Unterschiede subtil: Beide Seiten beschreiben sowohl Möglichkeiten für eine Konsistenz wie auch Inkonsistenz der geäußerten Schülervorstellungen. Wie diese sich jeweils empirisch zeigen, ist häufig abhängig von der Ebene der Betrachtung (feiner vs. grober Zerlegungsgrad der Vorstellungen) sowie der Operationalisierung von Konsistenz (diSessa, 2006). Dementsprechend ist auch die Befundlage uneinheitlich und beide Positionen können empirische Belege für ihre jeweilige Sichtweise anführen.

Eine Möglichkeit empirisch zwischen den Ansätzen zu unterscheiden, liegt in der Untersuchung von Personengruppen mit unterschiedlichen Antwortprofilen bei der Bearbeitung mehrerer Testaufgaben (z.B. Straatemeier, van der Mass & Jansen, 2008). So konnten z.B. Kleickmann und Kollegen (2011) in einer Stichprobe von 1817 Grundschulern Gruppen mit konsistentem und inkonsistentem Antwortverhalten auf Aufgaben zu den Themenbereichen Schwimmen/Sinken und Verdunstung/Kondensation identifizieren und werteten dies als Hinweis darauf, dass die verschiedenen Ansätze zur Beschreibung von initialem naturwissenschaftlichen Wissen nicht alternativ gesehen werden sollten, sondern zur Charakterisierung verschiedener Subpopulationen geeignet sind. Dabei stellten die Autoren (Kleickmann et al., 2011) fest, dass Kinder höherer Kompetenz eher in Gruppen mit konsistentem Antwortverhalten zu finden waren. Die Gruppenzugehörigkeit (konsistentes vs. inkonsistentes Antwortverhalten) war nur bedingt auf zuvor erhaltenen Unterricht in den in der Studie untersuchten Themenbereich zurückzuführen. Diese Befunde weisen damit in dieselbe Richtung wie die von Schneider und Hardy (2013), die mit der gleichen Methodik Schülerantworten aus einer längsschnittlichen Studie zum Thema Schwimmen/Sinken analysierten. Auch in dieser Studie konnten Schülergruppen mit unterschiedlich konsistenten Antwortmustern gefunden werden, die zur Vorhersage des weiteren Entwicklungsverlaufs (Abnahme, Stabilität oder Zunahme der Konsistenz) genutzt werden konnten (Schneider & Hardy, 2013). Eine eindeutige Interpretation der Befunde beider Studien im Hinblick auf den Kohärenz- oder Fragmentierungs-Ansatz ist nicht einfach möglich bzw. wäre eine umfassendere Theorie zur Erklärung der Befunde wünschenswert (diSessa, 2006).

Zur weiteren Möglichkeit einer empirischen Unterscheidung zwischen den beiden Positionen untersuchen Vertreter des Fragmentierungsansatzes häufig gezielt die Kontextgebundenheit von Schülervorstellungen und versuchen so, eine Inkonsistenz der Vorstellungen über verschiedene Kontexte hinweg zu zeigen (diSessa, 2006; diSessa et al., 2004). Auf die besondere Bedeutung des Kontextes bei der Erforschung von Schülervorstellungen wird im nächsten Abschnitt eingegangen.

### **2.2.3 Die Kontextualität von Schülervorstellungen**

Die Bedeutung des Kontextes für Schülervorstellungen ist in der Conceptual Change-Forschung eine viel diskutierte Frage. Oft wurde eine kontextabhängige Präferenz für eine Erklärung berichtet, die allerdings nicht mit einem tatsächlichen Conceptual Change verwechselt werden darf (Treagust & Duit, 2008). Dabei wurde jedoch uneinheitlich beschrieben, welche Aspekte einer Situation der Begriff Kontext umfasst und wie diese Aspekte auf die Vorstellungen wirken. So wurde von einigen Autoren die Nutzung von Konzepten im Alltag, in der Wissenschaft, in der Schule und in anderen unterschiedlichen Umgebungen als Kontextgebundenheit beschrieben (z.B. Larsson & Halldén, 2009; Stark, 2003), während von anderen bestimmte Merkmale von zu erklärenden Phänomenen oder konkreten Situationen als Kontextmerkmale herangezogen wurden (z.B. diSessa et al., 2004; Engel Clough & Driver, 1986).

Einigkeit herrscht in der Annahme, dass der Kontext als kritisches Element der Schülervorstellungen zu berücksichtigen ist. Besonders deutlich wird dies in Ansätzen der situierten Kognition, in denen die Autoren davon ausgehen, dass die Bedeutung von Konzepten kontextgebunden ist und dass Kontextbedingungen stark die Entwicklung von Konzepten beeinflussen (z.B. Anderson, Reder & Simon, 1996; Stark, 2003). Andere Autoren gehen davon aus, dass es sich bei der Veränderung von Schülervorstellungen um eine radikale Rekonstruktion des Kontexts handelt (Halldén, Scheja & Haglung, 2008; Larsson & Halldén, 2009).

Der Einfluss des Kontextes auf geäußerte Vorstellungen kann erklärt werden durch assoziative Vorgänge, die bei der Verarbeitung und Interpretation von Aufgabenkontextens stattfinden: neue Situationen werden zunächst mit Hilfe des eigenen Vorwissens bzw. eigener Erfahrungen mit ähnlichen Situationen interpretiert und bewertet (Driver & Easley, 1978; diSessa, 2006). Da vorunterrichtliche Konzepte eben geprägt sind durch diese eigenen Erfahrungen (z.B. Engel Clough & Driver, 1985) und diese jeweils in spezifischen Situationen, mit jeweils spezifischen Merkmalen, gemacht wurden, können durch bestimmte Aspekte dieser Situationen auch jeweils bestimmte Konzepte aktiviert werden (Tytler, 1998), die in anderen Fällen nicht mit dem gerade vorliegenden Kontext assoziiert sind.

Im extremsten Falle wurden Berichte über inkonsistente Schülerantworten auf verschiedene Aufgabenstellungen als reine Artefakte der eingesetzten Methode kritisiert (Engel Clough & Driver, 1986). Um zu validen Aussagen über Schülervorstellungen zu kommen, sei Konsistenz eine notwendige Voraussetzung. Da aber die Kontextgebundenheit u.U. ein zentrales Merkmal naiver naturwissenschaftlicher Vorstellungen ist und gerade eine stärkere Generalisierung dieser Vorstellungen über unterschiedliche Kontexte hinweg das Ziel naturwissenschaftlichen Lernens ist, erscheint es sinnvoll, die Einflüsse unterschiedlicher Kontexte gezielt zu untersuchen. Die so gewonnenen Erkenntnisse könnten sowohl für die Konstruktion von Aufgaben für die Erfassung von Schülervorstellungen wie auch für die Gestaltung von Lerngelegenheiten genutzt werden.

In den oben beschriebenen Theorien zu Kohärenz oder Fragmentierung (vgl. Kap. 2.2.2) wird der Einfluss des Kontextes auf die Schülervorstellungen jeweils unterschiedlich beschrieben. Ausgehend von den Ansätzen, in denen Schülervorstellungen ähnlich zu wissenschaftlichen Theorien konzeptualisiert werden, ist anzunehmen, dass die naiven Theorien einen vergleichbar breiten Anwendungsbereich aufweisen. Der Ansatz der Fragmentierung hingegen geht von einer kontextspezifischen Aktivierung der Wissens Elemente aus. Die Anwendungsbereiche der naiven Ideen sind dementsprechend schmaler und die Grenzen ihrer Anwendbarkeit können von der fachwissenschaftlichen Kategorisierung abweichen (diSessa et al., 2008), d.h. auch Phänomene, zu deren Erklärung aus fachlicher Sicht die gleichen Konzepte heran gezogen werden, werden u.U. von den Schülern unterschiedlich wahrgenommen und mit Hilfe unterschiedlicher Konzepte erklärt.

Entsprechend dieser unterschiedlichen Vorhersagen zeigen Vertreter des Rahmentheorieansatzes in Studien die logische Konsistenz und Kohärenz von Schülervorstellungen über Kontexte hinweg auf und stellen diese in (mentalen) Modellen dar (z.B. Vosniadou & Brewer, 1992). In einer Untersuchung an 105 Kindern zwischen 4 und 15 Jahren z.B. konnten Ioannides und Vosniadou (2002) zeigen, dass nur eine kleine Auswahl gut definierter und in sich konsistenter Interpretationen zur Erklärung des Wirkens physikalischer Kraft in unterschiedlichen Situationen herangezogen wurde. Auch Vosniadou und Brewer (1992) konnten in ihrer Untersuchung über Vorstellungen zur Erde die geäußerten Antworten auf wenige subjektive Theorien zurückführen.

Entgegen dieser Befunde konnte in vielen Studien bereits eine starke Kontextgebundenheit von Schülervorstellungen belegt werden (vgl. z.B. Caravita & Halldén, 1994). So zeigten z.B. diSessa und Kollegen (2004) in ihrer Studie, dass die Antworten von 30 Schülern auf Aufgaben zum physikalischen Kraft-Konzept in Abhängigkeit von Merkmalen der jeweiligen Aufgabenstellung variierten. Ebenso unterschieden sich bei Engel Clough und Driver (1986) die Schülerantworten auf die Frage nach evolutionärer Anpassung von Organismen an ihre Umwelt je nachdem, ob sich die Frage auf Raupen oder Polarfüchse bezog. In der gleichen Studie allerdings konnten hinsichtlich anderer Themenbereiche, wie z.B. dem Druck in Flüssigkeiten, keine Unterschiede im Hinblick auf unterschiedliche Aufgabenkontexte festgestellt werden (Engel Clough & Driver, 1986). Insgesamt beobachteten die Autoren, dass wissenschaftsnahe Erklärungen konsistenter von den Probanden genutzt wurden als alternative Erklärungen (Engel Clough & Driver, 1986).

Auch in Bezug auf die in der vorliegenden Arbeit untersuchten Themenbereiche Verdunstung/Kondensation sowie Schwimmen/Sinken wurden empirische Befunde zur Kontextualität von Schülervorstellungen berichtet; diese werden in Kapitel 2.2.4.2 und 2.2.5.2 dargestellt, nachdem in den jeweils voran stehenden Abschnitten (Kapitel 2.2.4.1 und 2.2.5.1) allgemein die Befunde zu Schülervorstellungen in diesen Inhaltsbereichen dargestellt werden.

## **2.2.4 Untersuchungen zu Schülervorstellungen bei Verdunstung und Kondensation**

### **2.2.4.1 Schülervorstellungen zu Verdunstung und Kondensation**

Der Themenbereich Verdunstung/Kondensation umfasst das Verständnis der Schüler von den Vorgängen beim Wechsel des Aggregatzustandes (hier von Wasser) von flüssig zu gasförmig und von gasförmig zu flüssig. Dazu gehört, neben einem Verständnis der Prozesse der Phasenübergänge selbst, ein Verständnis des Materialbegriffs und der fortgesetzten Existenz von Materie sowie des gasförmigen Zustands (Tytler, 2000).

Vor allem junge Kinder haben noch keine Vorstellung von der Erhaltung der Masse entwickelt und nehmen daher an, dass verdunstendes Wasser einfach verschwindet. So fanden Osborne und Cosgrove (1983) in ihrer Studie mit N=43 acht- bis siebzehnjährigen Schülern auf die Interviewfrage, was mit Wasser passiert, das von einer Untertasse verdunstet, über alle Altersgruppen hinweg die Antwort, dass das Wasser einfach trocknet und verschwindet. Die Antwort war allerdings bei jüngeren Schülern im Alter zwischen acht und zehn Jahren stärker verbreitet als in den anderen Altersgruppen und wurde mit zunehmendem Alter durch andere Vorstellungen abgelöst. Diesen Befund konnten die Autoren mit einem geschlossenen schriftlichen Aufgabentyp in einer Studie mit N=525 Schülern im Alter von zwölf bis siebzehn Jahren bestätigen (Osborne & Cosgrove, 1983). Schülervorstellungen dieser Art wurden auch in weiteren Studien bestätigt (z.B. Bar, 1989; Bar & Galili, 1994; Bar & Travis, 1991; Russell, Harlen & Watt 1989). Tytler (2000) konnte hingegen in seiner Studie mit Erst- und Sechstklässlern eine solche Nicht-Erhaltungsvorstellung nur bei wenigen Schülern belegen. Selbst Schüler, die das Verschwinden von Wasser in ihren Antworten thematisierten, taten dies häufig in Kombination mit anderen Vorstellungen. Der Autor fand hingegen Antworten, in denen das Wasser einfach als trocknend, ohne weitere Erklärung („just drying up“), beschrieben wurde. Seiner Ansicht nach ist die Interpretation der Schülerantworten im Sinne der Nicht-Erhaltung eine Interpretation aus Erwachsenensicht: Während besonders die jüngeren Schüler bei dieser Frage das für sie Wahrnehmbare beschreiben (Wasser ist nicht mehr sichtbar, somit aus dem Bereich des direkt Wahrnehmbaren verschwunden), ist für sie die Frage, was mit dem getrockneten Wasser passiert, nicht von Belang (Tytler, 2000).

Auch wenn bereits ein Verständnis der Massenerhaltung erreicht wurde bzw. aus den Antworten zu entnehmen ist, geht dies häufig einher mit anderen fehlerhaften Vorstellungen. Dazu gehört die Vorstellung, dass das verdunstende Wasser in die Gegenstände bzw. Oberflächen einzieht oder einsickert (z.B. Beveridge, 1985; Osborne & Cosgrove, 1983; Tytler, 2000). Bei Russell und Kollegen (1989) gaben 19% der Acht- bis Neunjährigen und 21% der Zehn- bis Elfjährigen eine entsprechende Vorstellung auf die Frage hin, was mit dem Wasser beim Trocknen von nassen Kleidungsstücken passiert. Ging es hingegen um das Absinken des Wasserspiegels in einem Wassertank, wurde diese Vorstellung nicht geäußert (Russell et al., 1989). Diese Vorstellung war,

wie die Nicht-Erhaltungsvorstellung, mit zunehmendem Alter seltener zu finden (Bar & Galili, 1994; Osborne & Cosgrove, 1983). Eine weitere fehlerhafte Interpretation der Verdunstung besteht in der Annahme, dass ein aktiver Agent (z.B. eine Mensch, ein Tier, die Sonne) das Wasser zu einem anderen Ort bewege (Russell et al., 1989). Neben den bisher beschriebenen naiven Vorstellungen, die den Verdunstungsprozess auf einer makroskopischen Ebene thematisieren, findet sich außerdem die Vorstellung, dass Wasser sich beim Verdunsten in Wasserstoff und Sauerstoff aufspaltet (Osborne & Cosgrove, 1983).

Mit zunehmendem Alter werden die fehlerhaften Konzepte durch die Vorstellung ersetzt, dass das Wasser in die Luft, die Wolken oder den Himmel aufsteigt, ohne dass eine Veränderung des Aggregatzustandes mit dem Aufsteigen assoziiert wird. Bei Russell und Kollegen (1989) war diese Antwort die häufigste auf beide gestellten Aufgaben (sinkender Spiegel in einem Wassertank, trocknende Wäsche) unter den Acht- bis Neun- (33% bzw. 50%) und den Zehn- bis Elfjährigen (57% bzw. 70%), während diese Vorstellung bei den jüngeren Kindern nicht bzw. nur selten vertreten war. Diese fortgeschrittene Vorstellung war allerdings in anderen Studien deutlich seltener zu finden, beispielsweise nur bei 4% der Zwölfjährigen bei Osborne & Cosgrove (1983) oder bei 6% bzw. 8% der Sechs- oder Achtjährigen bei Tytler (2000), jeweils auf die Frage, was mit dem trocknenden Wasser auf einer Untertasse passiert.

Ein Verständnis von Verdunstung, dass sowohl den Ortswechsel sowie eine physikalische Veränderung des Wassers (Wechsel des Aggregatzustandes) enthält, wird von den jüngeren Schülern häufig noch nicht erreicht, da damit eine Akzeptanz der Vorstellung, dass sich ständig nicht-sichtbares, gasförmiges Wasser (Wasserdampf) in der Luft befindet, einhergeht (Tytler, 2000). Bei Russell und Kollegen (1989) erreichten die Fünf- bis Siebenjährigen diese Vorstellung noch nicht, während sie von den älteren Kindern bereits mit 17% bei den Acht- bis Neujährigen und 9% bei den Zehn- bis Elfjährigen bei der Erklärung des sinkenden Wasserspiegels in einem Behälter vertreten wurde. Allerdings waren entsprechende Antworten auch für diese Altersgruppen nicht auf die Frage, was mit dem Wasser aus trocknender Wäsche passiert, zu finden (Russell et al., 1989).

Die beschriebenen Befunde beziehen sich alle auf die Erklärung von Verdunstungs-Phänomenen unterhalb des Siedepunktes. In verschiedenen Studien konnte gezeigt werden, dass ein wissenschaftlich adäquates Verständnis von Verdampfung, im Sinne der korrekten Beschreibung aufsteigenden Wassers in veränderter Form, bei Situationen oberhalb des Siedepunktes von den Schülern schon früher erreicht werden kann (z.B. Bar & Galili, 1994; Bar & Travis, 1991; Johnson, 1998; Tytler, 2000). Die Sichtbarkeit von aufsteigenden, kondensierten Wassertröpfchen beim Kochen, umgangssprachlich als Wasserdampf bezeichnet, erleichtert vermutlich die korrekte Interpretation des Phänomens. Allerdings argumentiert Tytler (2000), dass der aufsteigende, sichtbare Dampf von den Probanden seiner Studie häufig noch nicht mit dem sinkenden Wasserspiegel oder den im kochenden Wasser aufsteigenden Blasen korrekt in Verbindung

gebracht werden kann. Zudem ist in diesem Zusammenhang eine weitere Fehlvorstellung festzustellen: Einige Schüler nehmen an, dass Verdampfung nur bei kochender Flüssigkeit oder expliziter Energiezufuhr stattfinden kann (Costu & Ayas, 2005).

Ein angemessenes Verständnis von Kondensation ist in der Regel erst nach dem Verständnis des Verdunstungsvorgangs zu erreichen, da hierfür zunächst begriffen werden muss, dass in der Luft ständig Wasserdampf vorhanden ist (Johnson, 1998; Tytler, 2000). Auch bei diesem Vorgang sind bei Schülern bis zum Erreichen wissenschaftlich adäquater viele nicht-wissenschaftliche Vorstellungen zu finden. Im Gegensatz zu Verdunstung greifen Schüler zur Erklärung von Kondensation häufiger auf eigene, anekdotische Erfahrung zurück oder bilden Analogien (Tytler, 2000) oder sie können, wenn sie bei der Erklärung von Verdunstung den Wechsel des Aggregatzustandes noch nicht berücksichtigen, größtenteils keine Erklärung liefern (Johnson, 1998).

Erste Erklärungsansätze bei Aufgaben, bei denen Kondensation an kalten Gegenständen erklärt werden soll, thematisieren häufig, dass das Wasser durch Materie hindurch, z.B. durch die Wände eines Gefäßes mit Wasser, nach außen kommt (vgl. Osborne & Cosgrove, 1983; Tytler, Peterson & Prain, 2006). Häufig werden aber auch Varianten dieser Vorstellung berichtet, bei der Kälte durch die Materie nach außen kommt und Wassertropfchen bildet (Driver, Squires, Rushworth & Wood-Robinson, 1994; Osborne & Cosgrove, 1983; Tytler et al., 2006). Dieses Antwortmuster verdeutlicht, dass die befragten Schüler noch nicht klar zwischen Dingen und Zuständen differenzieren können. Eine weitere häufige Vorstellung, die mit Kondensation an kalten Gegenständen verbunden ist, ist die Idee, dass eine dünne Eisschicht auf dem Gegenstand schmilzt und sich deswegen Wassertropfen bilden (Johnson, 1998). Analog zu der Vorstellung, dass Wasser sich während des Verdunstens in Wasserstoff und Sauerstoff aufspaltet, äußern Schüler die Vorstellung, dass Wasserstoff und Sauerstoff zusammen Wassertropfen bilden (Driver et al., 1994, Osborne & Cosgrove, 1983). Bei Aufgaben, die das Niederschlagen von sichtbarem Wasser“dampf“ an Gegenständen thematisieren (z.B. ein Teller wird über einen Topf mit kochendem Wasser gehalten), kommen außerdem noch die Vorstellungen, der Gegenstand schwitze (Osborne & Cosgrove, 1983) oder die Verwandlung von Hitze in Wasser, zum Tragen. Erst nach und nach können die Schüler die Rolle der umgebenden Luft und des darin vorliegenden gasförmigen Wassers für die Kondensation verstehen (Bar & Travis, 1991). Ein Verständnis, das sowohl den Ortswechsel wie auch den Wechsel des Aggregatzustandes berücksichtigt, wird auch bei Kondensation nur von älteren Schülergruppen erreicht.

#### *2.2.4.2 Kontextualität von Schülervorstellungen zu Verdunstung und Kondensation*

In vielen Studien, die sich mit Schülervorstellungen zum Thema Verdunstung und Kondensation beschäftigten, konnte festgestellt werden, dass sich die geäußerten Konzepte in Abhängigkeit von der Aufgabenstellung unterschieden. Beispielsweise konnten Russell und Kollegen (1989) bei den

von ihnen untersuchten Aufgabenstellungen eine höhere Prävalenz der Vorstellung „Wasser wird aufgesaugt/zieht ein“ für die Aufgabenstellung „Wäsche trocknet“ feststellen als bei der Aufgabenstellung „fallender Wasserspiegel in einem Wassertank“. Entsprechend fanden Bar und Galili (1994) diese Vorstellung bei dem Trocknen von Wäsche häufiger als beim Trocknen von verschüttetem Wasser auf dem Boden, während dafür bei der letztgenannten Situation die Vorstellung „Wasser steigt in die Luft“ häufiger geäußert wurde.

Costu und Ayas (2005) untersuchten in ihrer Studie mit N=313 Sekundarschülern gezielt, inwiefern sich die geäußerten Vorstellungen in Abhängigkeit von verschiedenen, verdunstenden Flüssigkeiten unterschieden. Sie führten dazu einen schriftlichen Test mit offenem Antwortformat durch in dem die Aufgabenstellungen sich auf „Wasser“, „Alkohol“ oder „eine Flüssigkeit“ bezogen. Tatsächlich stellten sie fest, dass einige Vorstellungen nur in Bezug auf bestimmte Flüssigkeiten gehalten wurden (Costu & Ayas, 2005). So äußerten die Schüler z.B. die Vorstellung, dass nur Wasser aus dem Alkohol verdunste oder dass bei Alkohol die Temperatur einen stärkeren Einfluss auf die Verdunstungsrate habe. Tytler (2000) beschreibt, dass in Abhängigkeit von der Temperatur bzw. der Präsenz einer Wärmequelle, des physikalischen Ortes (drinnen oder draußen) und anderen Umständen unterschiedliche Vorstellungen provoziert werden. Er erklärt dies damit, dass Assoziationen den Kern des Denkens bilden (vgl. diSessa, 1993), die sich entsprechend bei den Situationen unterscheiden.

## **2.2.5 Untersuchungen zu Schülervorstellungen zum Schwimmen und Sinken**

### **2.2.5.1 Schülervorstellungen zum Schwimmen und Sinken**

Zum Verständnis von Schwimmen/Sinken gehören ein Verständnis der Auftriebskraft von Wasser sowie ein Verständnis der Dichte bzw. des Dichtevergleichs von Objekten mit der umgebenden Flüssigkeit, der Verdrängung und der Abhängigkeit der Verdrängung vom Volumen eines Objektes.

Die Auftriebskraft, die eine wichtige Kraftwirkung in der Erklärung des Schwimmens oder Sinkens darstellt, wird ausgeübt durch die das Objekt umgebende Flüssigkeit, welche i.d.R. Wasser ist. Diese zentrale Rolle des Wassers für das Schwimmverhalten wird bei frühen Erklärungen häufig noch nicht berücksichtigt. Junge Kindern äußern animistische oder moralische Erklärungen (vgl. Piaget, 2001) oder auch einen Vergleich mit bereits gemachten Erfahrungen oder den Erklärungen, die sie von Erwachsenen erhalten haben (Butts, Hoffmann & Anderson, 1993). Wenn erste eigenständige Erklärungen generiert werden, beziehen diese sich häufig auf einzelne Attribute konkreter Objekte (Butts et al., 1993; Piaget, 2001). So argumentierten bei Hsin und Wu (2011) die Befragten z.B. mit der Größe, dem Gewicht oder dem Material, wobei von den Erklärungen der fünfjährigen Probanden für Schwimmen und Sinken über die Hälfte auf Argumentationen mit dem Gewicht entfielen, während bei den Vierjährigen nur 30% der Erklärung darauf entfielen. Auch in Biddulph und Osbornes Studie (1984), in der sie sieben- bis vierzehnjährige Schüler nach ihren

Erklärungen für Schwimmen bzw. Sinken fragten, wurde das Gewicht bei den Probanden, die nur einen singulären Grund für das Schwimmverhalten angaben, ungeachtet der betrachteten Altersgruppe, häufig genannt. Interessant in Bezug auf diese Gewichtsvorstellung ist dabei, dass jüngere Kinder etwa im Alter von fünf bis sechs Jahren erklären, dass Gegenstände schwimmen, weil sie schwer sind (Piaget, 2001) und mit diesem Gewicht ausreichende Kraft mitbringen, sich selbst über Wasser zu halten. Mit zunehmendem Alter, etwa im Alter von sechs bis acht Jahren, kehrt sich dieses Muster um in die Vorstellung, dass leichte Dinge schwimmen und schwere untergehen (Buck, 2008; Piaget, 2001). Ein anderes Attribut, das zur Erklärung von Schwimmen oder Sinken von Objekten herangezogen wird, ist die Form von Objekten, wobei häufig darauf fokussiert wird, ob Dinge eine hohle Form haben bzw. Luft in ihnen eingeschlossen ist (Engelen, Jonen & Möller, 2002; Fischer, 2008; Pollmeier et al., 2011). Viele der von Probanden herangezogenen Attribute könnten, würden Sie zusammengebracht mit anderen relevanten Eigenschaften, nützlich für die Erklärung des Schwimmverhaltens sein (Havu-Nuutinen, 2005). Zudem fällt bei der Untersuchung der Schülervorstellungen zu diesem Themenbereich besonders auf, dass häufig mehrere Gründe für das Schwimmen oder Sinken genannt werden (Biddulph & Osborne, 1984; Engelen et al., 2002; Pollmeier et al., 2011), oder dass für unterschiedliche Objekte immer wieder neue Erklärungen generiert werden (Biddulph & Osborne, 1984; Tytler & Peterson, 2004).

Ein beginnendes Verständnis von Dichte ist nach Piaget (2001) erst ab einem Alter von sechs bis acht Jahren zu erreichen. Allerdings zeigten neuere Studien, dass ein solches Verständnis auch bei jüngeren Kindern möglich ist. So belegten Smith, Carey und Wiser (1985), dass auch jüngere Kinder Aufgaben lösen konnten, die ein qualitatives Verständnis von Dichte erfordern. In ihrer Interviewstudie mit drei- bis neunjährigen Kindern fragten sie die Probanden danach, welcher von zwei Gegenständen aus dem schwereren Material gemacht ist. Während die Drei- bis Vierjährigen meistens noch allein nach dem Gewicht urteilten, ließen bereits die Fünf- bis Siebenjährigen ein vorhandenes, wenn auch noch undifferenziertes, Dichte-Konzept erkennen. In Bezug auf die Dichte von Flüssigkeiten sind allerdings noch weitere Fehlvorstellungen zu finden. So fanden Fassouloupoulos, Kariatoglou und Koumaras (2003) in einer Fragebogenstudie mit dreihundert 12 bis 15 Jahre alten Kindern die Vorstellungen, dass die Dichte mit der Menge einer Flüssigkeit ansteigt oder in kleineren Gefäßen zunimmt. Analog dazu fanden sich Vorstellungen, dass der Druck (in einer Flüssigkeit) ebenfalls durch ihre Menge bestimmt sei, mit der Tiefe zu- bzw. abnehme oder von den Seiten des Gefäßes aufgebaut wird. Diese Vorstellungen waren aber in Bezug auf die Dichte seltener zu finden als in Bezug auf den Druck, der von den Flüssigkeiten ausgeübt wird (Fassouloupoulos et al., 2003).

Bei dem Verständnis von Verdrängung geht es besonders darum zu erkennen, dass die Menge des verdrängten Wassers vom Volumen eines eingetauchten Gegenstands abhängt. In der Studie von Biddulph und Osborne (1984) äußerten 53 von 80 Schülern zwischen sieben und vierzehn Jahren,



dass sich die Menge des verdrängten Wassers mit einer Verformung des eingetauchten Gegenstandes verändert. Allerdings nahm die Mehrheit von ihnen an, die Verdrängung nehme ab. Hingegen konnten nur 11 der 80 Schüler richtig beantworten, dass die Verdrängung zunimmt, wenn der Gegenstand so verformt wird, dass es schwimmt. Neben der Form wird außerdem häufig auf das Gewicht der Gegenstände fokussiert und die Schüler nehmen an, die Menge des verdrängten Wassers sei bei schweren Gegenständen größer (Pollmeier et al., 2011).

Neben den Erklärungen, die für das Schwimmverhalten gegeben werden, untersuchten Biddulph und Osborne (1984) in ihrer Studie mit 104 sieben- bis vierzehnjährigen Schülern, welche Situationen von den Probanden überhaupt als „Schwimmen“ angesehen werden. Es stellte sich dabei heraus, dass lediglich Objekte, die zum Teil ins Wasser eingetaucht, größtenteils aber oberhalb der Wasseroberfläche waren, von der Mehrheit der Probanden als schwimmend eingeschätzt wurden, wohingegen, Objekte, deren größter Teil unter der Wasseroberfläche lag, als teils schwimmen, teils sinkend angesehen wurden. Sowohl Objekte, die komplett auf der Wasseroberfläche waren oder die komplett unter Wasser, aber frei schwebend, waren, wurden nicht als schwimmend angesehen.

Diese Befunde decken sich nur zum Teil mit denen von Joung (2009), der in seiner Studie 148 Fünftklässler Situationen zeichnen ließ, die sie spontan mit den Begriffen Schwimmen bzw. Sinken assoziierten. Die Mehrheit der Schüler zeichnete, gefragt nach einer „Situation, in der ein Ding im Wasser schwimmt“, ein Objekt auf dem Wasser (48.0%) oder halbeingetaucht (30.4%). Gefragt nach einem sinkenden Objekt antwortete die Mehrheit der Schüler mit einem Objekt ganz am Boden (33.8%) oder auf halbem Weg zwischen Wasseroberfläche und Grund (28.4%). Neben der Position des Objekts analysierte Joung (2009) außerdem den räumlichen Hintergrund sowie die Art des Objektes, die die Probanden in ihren Zeichnungen wählten. Dabei zeigte sich, dass sowohl bei den Situationen zum Schwimmen wie auch bei denen zum Sinken mehrheitlich große Gewässer, wie das Meer oder ein Fluss, gewählt wurden. Das gewählte Objekt war in beiden Fällen häufig ein Mensch. Auffällig war aber, dass beim Sinken fast doppelt so häufig ein Boot gezeichnet wurde (21.6%) wie beim Schwimmen (12.2%). Dieser letzte Befund zeigt, dass v.a. das Sinken dynamisch betrachtet wird, mit einem Boot, das sich im Prozess des Untergehens befindet, aber kein Gegenstand ist, der üblicherweise auf den Grund des Wassers sinkt.

#### *2.2.5.2 Kontextualität von Schülervorstellungen zum Schwimmen und Sinken*

In Studien zu Verdunstung und Kondensation werden v.a. Merkmale der umgebenden Situation, wie z.B. ob sie drinnen oder draußen situiert sind, als relevant für die Kontextualität von Schülervorstellungen beschrieben. Im Zusammenhang mit dem Thema Schwimmen und Sinken wurde hingegen davon berichtet, dass Eigenschaften des zentralen, betrachteten Objektes für die Antwortgenerierung herangezogen werden.

So berichteten Biddulph & Osborne (1984), dass in Ihrer Interviewstudie mit 144 Sieben- bis Vierzehnjährigen nur ein geringer Teil der Probanden für das Schwimmverhalten verschiedener Objekte wiederholt die gleiche Erklärung verwendete. Die Mehrheit gab für unterschiedliche Gegenstände jeweils unterschiedliche Erklärungen. Diese Befunde decken sich mit denen von Hsin und Wu (2011), die ein entsprechendes Antwortmuster in Interviews mit deutlich jüngeren Kindern feststellen konnten, deren Antworten sich je Objekt auf dessen spezifische Eigenschaften, wie das Gewicht, die Größe oder die Härte bezogen. Dies führt, da nicht alle Gegenstände die gleichen Merkmale teilen, über verschiedene Aufgaben hinweg betrachtet zu einem inkonsistenten, kontextuell variablen Antwortverhalten.

Vergleichbar wird in Untersuchungen zum Dichteverständnis davon berichtet, dass spezifische Charakteristika von Gegenständen die Urteile über deren relative Dichte beeinflussten (Smith et al., 1985) und dass ein gezeigtes inkonsistentes Antwortmuster von den Merkmalen der Aufgaben abhing (Fassoulopoulos et al., 2003).

Ein umgekehrter Zugang zu der Frage nach der Kontextualität ist in den Studien zu finden, in denen die Schüler gebeten wurden zu beschreiben oder zeichnen, welche Situation sie sich als Schwimmen bzw. Sinken vorstellen (Biddulph & Osborne, 1984; Joung, 2009). Obwohl aus fachlicher Sicht zu Erklärung in beiden Fällen die gleichen Konzepte genutzt würden, subsumierten die Probanden beider Studien oberflächlich unterschiedliche Phänomene unter den Begriffen. Daher ist zu vermuten, dass entsprechend unterschiedlich zur Erklärung der Phänomene argumentiert würde.

### **2.2.6 *Conceptual Change, Kompetenzmodelle und Kontexteinfluss***

Die Theorie der Konzeptveränderung bzw. des Conceptual Change (siehe Kapitel 2.2.1) ist eine zentrale Lerntheorie in der Fachdidaktik für Naturwissenschaften in der Grundschule: Sie geht davon aus, dass Schüler bereits mit vielfältigen, aber teils defizitären Vorstellungen in den naturwissenschaftlichen Unterricht kommen, welche durch konstruktives Lernen in Richtung adäquater Vorstellungen verändert werden sollen. Noch offene Fragen bestehen in dieser Forschungsrichtung, z.B. hinsichtlich der Struktur dieses initialen naturwissenschaftlichen Wissens, wobei zwischen Ansätzen, die eine theorieähnliche Konsistenz der Vorstellungen postulieren, und Ansätzen, die von einer stärkeren Fragmentierung des Wissens ausgehen, unterschieden wird (siehe Kapitel 2.2.2). Ein zentraler Faktor bei der Untersuchung von Schülervorstellungen ist der (Anwendungs-) Kontext dieser Vorstellungen. Verschiedene empirische Studien konnten einen Einfluss des Kontexts belegen. Dabei wurde einerseits die Plausibilität der Ansätze zu Kohärenz oder Fragmentierung zur Erklärung von Antwortmustern auf mehrere Aufgabenkontexte untersucht (siehe Kapitel 2.2.3). Andererseits wurden für die in dieser Arbeit untersuchten Themenbereiche explizit Befunde zur Kontextabhängigkeit der Vorstellungen in diesen Bereichen berichtet (siehe Kapitel 2.2.4.2 und 2.2.5.2).

Auf die Frage nach dem Einfluss des Kontexts auf die untersuchten Kompetenzen wird in den in Kapitel 2.1.2 vorgestellten Kompetenzmodellen noch nicht ausreichend eingegangen. Mit Hinweis auf die Kontextualität von Schülervorstellungen wird im Modell von Schecker und Parchmann (2006) zwar eine entsprechende Dimension berücksichtigt. Eine vergleichbare Unterscheidung wird in den anderen Modellen nicht vorgenommen. Dabei wird in Bezug auf den Kompetenzbegriff gerade der Bezug auf konkrete Anforderungssituationen, also die Kontextabhängigkeit, als wesentliches Merkmal herausgestellt (Klieme & Leutner, 2006; Weinert, 2001; 2002) und die Erfassung und Überprüfung entsprechender Merkmale gefordert (Klieme & Leutner, 2006). Eine Modellierung naturwissenschaftlicher Kompetenzen, v.a. mit Rückgriff auf die Forschung zu Schülervorstellungen und Konzeptwechsel, sollte also auf eine Erfassung dieser Kompetenzen in einer möglichst breiten Auswahl konkreter Anforderungssituationen setzen und den Einfluss von Kontextmerkmalen auf die gezeigten Kompetenzen untersuchen. Dabei können die Kontextmerkmale als Merkmale der Aufgaben, die zur Erfassung eingesetzt werden, verstanden werden. Das folgende Kapitel widmet sich daher der Rolle von Aufgabenmerkmalen in Bezug auf Kompetenzmodellierung sowie der Bedeutung von Kontextmerkmalen als einer möglichen Kategorie von Aufgabenmerkmalen.

## **2.3 Aufgabenmerkmale**

### ***2.3.1 Die Rolle von Aufgabenmerkmalen bei der Validierung von Kompetenzmodellen***

Die Überprüfung der Gültigkeit von Kompetenzmodellen (siehe Kapitel 2.1) erfordert zunächst die Umsetzung in Testaufgaben, die den unterschiedlichen Zellen bzw. Bereichen des Modells zugeordnet werden können. Im Falle von angenommenen Kompetenzniveaus werden i.d.R. Aufgaben mit bestimmten Merkmalen realisiert, deren erfolgreiche Bearbeitung die für ein Niveau charakteristischen Fähigkeiten erfordert. Daher kommt den Aufgabenschwierigkeiten, verstanden als empirisch ermittelte Lösungshäufigkeiten als Resultat einer Interaktion zwischen Aufgaben und Probanden, eine entscheidende Rolle bei der Validierung der Aufgaben zu (vgl. z.B. Kauertz, 2008; Hartig & Frey, 2012): Aufgaben, die eine Kompetenzausprägung im unteren Bereich erfassen, sollten häufiger gelöst werden als Aufgaben, die eine Ausprägung im oberen Bereich erfassen. Sollte sich also die vorab angenommene Schwierigkeitsstufung der Aufgaben mit der empirisch ermittelten decken, wäre das der Aufgabenentwicklung zugrunde liegende Modell zur Erklärung der unterschiedlichen Fähigkeitsausprägungen der Testpersonen gut geeignet. Die mit den Niveaus verbundenen Aufgabenmerkmale können zur Quantifizierung dieses Zusammenhangs genutzt werden. Diese Form der Validierung wird als bedeutender Teil der Konstruktvalidierung gesehen (Embretson, 1983; Hartig & Frey, 2012; Kauertz, 2008), da aus der theoretisch abgeleiteten Operationalisierung des untersuchten Konstrukts plausible Hypothesen über die Schwierigkeitsstufung der Aufgaben abgeleitet und überprüft werden können (Rost, 1996). Ein

solches Vorgehen hatte ursprünglich v.a. in der kognitiven Psychologie Relevanz, da so die an der Aufgabenlösung beteiligten kognitiven Prozesse identifiziert werden sollten (Embretson, 1983). Im Zusammenhang mit Kompetenzmodellierung werden häufig eher deskriptive Aufgabenmerkmale untersucht, was dennoch zu wertvollen Aussagen führen kann. Neben der Validierung können die Aufgabenmerkmale auch zur kriteriumsorientierten Beschreibung von Skalen sowie, nachdem die Beiträge zur Schwierigkeit einzelner Merkmale quantifiziert wurden, für die Neuentwicklung von Aufgaben genutzt werden (Hartig & Frey, 2012).

Kauertz (2008) hat für die Rolle von Aufgabenmerkmalen ein Rahmenmodell zur Validierung von Kompetenzmodellen formuliert: Die tatsächliche, wahre Kompetenz, deren Messung intendiert ist, wird umgesetzt in eine theoretische Kompetenz bzw. Operationalisierung. Während die wahre Kompetenz die Leistung der Personen beeinflusst, wirkt die theoretische Kompetenz auf die Aufgabenkonstruktion, bei der bestimmte Merkmale realisiert werden. Aus der Interaktion der Personen bzw. ihrer Fähigkeiten mit den Aufgaben, die bestimmte Merkmale aufweisen, ergibt sich die empirische Aufgabenschwierigkeit. Für alle Schnittstellen zwischen den Elementen des Modells bestehen Möglichkeiten der Überprüfung bzw. Quantifizierung. Die Aufklärung der Schwierigkeitsvarianz durch Aufgabenmerkmale gibt Hinweise auf die Unschärfe des Modells an der Schnittstelle der Interaktion zwischen Personen und Aufgabenmerkmalen, die z.B. durch unterschiedliche Skalenniveaus von Aufgabenschwierigkeit und –merkmalen und durch das Wirken weiterer unberücksichtigter Merkmale entstehen kann (Kauertz, 2008).

Es existieren unterschiedliche Möglichkeiten zur genauen Analyse des Einflusses von Aufgabenmerkmalen auf die Schwierigkeit: Nach dem Einsatz des Instrumentes können die gewonnenen Daten analysiert und Schwierigkeitsindizes, beispielsweise mittels Rasch-Skalierung, ermittelt werden. Diese können dann im weiteren Verlauf selber zu abhängigen Variablen in den Analysen werden, und die Varianz der Schwierigkeitsindizes kann mit Hilfe varianz- oder regressionsanalytischer Verfahren aufgeklärt werden. Daneben existieren Modelle der probabilistischen Testtheorie, in denen die Skalierung der Daten unmittelbar mit der Analyse der Aufgabenmerkmale verbunden werden kann, wie z.B. das linear-logistische Testmodell (Fischer, 1973) oder das mehrdimensionale Komponentenmodell (Whiteley, 1980). Diese Vorgehensweise hat die Vorteile, dass der Einfluss der untersuchten Merkmale auf einzelne Aufgaben mit einem Koeffizienten quantifiziert werden kann, dass simultan Messwerte für Personen und Aufgaben geschätzt werden können und dass durch die Abbildung von Personen und Aufgaben auf einer gemeinsamen Skala die Interpretation von Fähigkeiten direkt mit bestimmten Aufgabenmerkmalen in Verbindung gebracht werden kann (Embretson, 1983).

Neben Aufgabenmerkmalen, die direkt mit der Operationalisierung der untersuchten Kompetenz im Zusammenhang stehen, können parallel weitere konstruktirrelevante Merkmale und ihr Einfluss auf die Aufgabenschwierigkeit im Sinne von Kontrollvariablen untersucht werden (Hartig & Frey, 2012). Die Erklärungsmacht von inhaltlich relevanten und irrelevanten Merkmalen kann

miteinander verglichen werden, um so das Ausmaß des jeweiligen Erklärungspotentials einschätzen zu können.

### **2.3.2 Unterschiedliche Arten von Aufgabenmerkmalen**

#### *2.3.2.1 Systematisierung von Aufgabenmerkmalen*

Nach dem Ausmaß der bei der Zuweisung von Aufgaben zu Merkmalen notwendigen Interpretation kann, in Entsprechung zur Unterscheidung bei der Beurteilung von Unterrichtsverhalten, zwischen niedrig- und hoch-inferenten Merkmalen unterschieden werden (Clausen, Reusser & Klieme, 2003; Kauertz, 2008; Rosenshine, 1970): Als niedrig-inferent werden Merkmale bezeichnet, die der direkten Beobachtung zugänglich und objektiv, d.h. mit einem geringen Maß an Interpretation, feststellbar sind. Hingegen werden Merkmale, die interpretative bzw. schlussfolgernde Prozesse erfordern und nicht direkt beobachtbar sind, als hoch-inferent bezeichnet. Diese Unterscheidung ist dabei nicht als strikte Dichotomie, sondern als Kontinuum zu verstehen.

Prenzel und Kollegen (Prenzel, Häußler, Rost & Senkbeil, 2002) haben in ihrer Analyse von PISA-Aufgaben zwischen drei Arten von Analysekatégorien für Aufgabenmerkmale unterschieden:

- Formale Aufgabenmerkmale: Hierunter verstehen die Autoren äußere Faktoren, die das Aussehen der Aufgaben und die Art der geforderten Reaktion bestimmen, z.B. die Länge des Aufgabentextes oder das Antwortformat.
- Kognitive Anforderungen beim Lösen der Aufgaben: Hierbei geht es um Vorgänge, von denen anzunehmen ist, dass sie bei der Aufgabenbearbeitung ablaufen, z.B. Verarbeitung von Textinformationen oder Abruf von Wissen aus dem Gedächtnis.
- Merkmale der für das Lösen der Aufgaben erforderlichen Wissensbasis: Hierzu gehören z.B. die Kategorien *Lösungsrelevante Informationen im Text* oder *Terminologisches Wissen*.

Als Weiterentwicklung dieser Systematik hat Kauertz (2008) drei Kategorien für die Analyse von Aufgabenmerkmalen vorgeschlagen, um dabei eine deutliche Trennung zwischen dem Inhalt der erwarteten Lösung und der kognitiven Verarbeitung zu schaffen: formale, inhaltliche und prozessorientierte Aufgabenmerkmale. Nach dem Ausmaß der bei der Zuweisung von Aufgaben zu den Merkmalen notwendigen Interpretation, werden dabei die ersten beiden Kategorien als niedrig-inferent und die letzte Kategorie als hoch-inferent erachtet. Im Folgenden werden diese drei Kategorien näher erläutert und ausgewählte Forschungsergebnisse berichtet.

#### *2.3.2.2 Formale Aufgabenmerkmale*

Zu den formalen Aufgabenmerkmalen gehören äußere Merkmale, wie der Umfang des Textes einer Aufgabe, das Antwortformat oder die Form und der Umfang von in der Aufgabe gegebenen Informationen (Kauertz, 2008). Bei der Konstruktion von Aufgaben sind diese Merkmale

zwangsläufig in den Aufgaben enthalten, obwohl sie inhaltlich nicht unbedingt im Zusammenhang mit dem theoretisch zugrundeliegenden Konstrukt stehen. Dennoch konnte auch konstruktirrelevanten Merkmalen oft ein bedeutsamer Einfluss auf die Aufgabenschwierigkeit nachgewiesen werden (Hartinger & Frey, 2012).

Beispielsweise konnte für mathematische und naturwissenschaftliche Aufgaben für den Einsatz am Ende der Sekundarstufe ein Einfluss des Grades der Offenheit des Antwortformates belegt werden (Draxler, 2006; Fischer & Draxler, 2006; Klieme, 2000). Auch Prenzel und Kollegen (Prenzel et al., 2002) konnten zeigen, dass die Erfordernis einer freien langen Antwort Aufgaben im PISA-Naturwissenschaftstest erschwerte, während eine Grafik oder bildliche Information das Lösen erleichterte. In einem physikalischen Leistungstest für die Sekundarstufe zeigte sich das Antwortformat gegenüber anderen formalen Aufgabenmerkmalen allerdings als vernachlässigbar, während die Position einer Aufgabe im Testheft und der Abstand zwischen dem Lern- und dem Testzeitpunkt mehr Schwierigkeitsvarianz aufklären konnten (Kauertz, 2008).

In einer Studie mit Grundschulern zum Themenbereich Aggregatzustände konnte ein unterschiedliches Antwortverhalten in Abhängigkeit vom vorgelegten Aufgabenformat (geschlossene vs. offene schriftliche Aufgaben) gezeigt werden (Bar & Travis, 1991). Auch in einer Validierungsstudie zum Projekt Science-P (siehe Kapitel 2.1.2.3), in der geschlossene schriftliche Aufgaben mit offenen Interviewaufgaben verglichen wurden, zeigte sich, dass die Interviewaufgaben für die befragten Drittklässler insgesamt schwieriger zu beantworten waren als die Testaufgaben, was vermutlich auf die Offenheit des Antwortformats zurück zu führen ist (Pollmeier et al., 2011).

#### 2.3.2.3 *Inhaltliche Aufgabenmerkmale*

Inhaltliche Aufgabenmerkmale beziehen sich, neben dem konkreten Inhalt einer Aufgabe, außerdem auf dessen Struktur. Die hohe Relevanz dieser Kategorie von Merkmalen ergibt sich aus ihrem direkten Bezug zur Operationalisierung des zu messenden Konstrukts.

Zur Beschreibung der Inhaltsstruktur physikalischer Leistungstestaufgaben wählte Kauertz (2008) die Merkmalsdimensionen *Komplexität*, *Leitidee einer Aufgabe* und *kognitive Aktivität* (siehe auch Kapitel 2.1.2.3) und konnte in seiner Arbeit deren Erklärungsmacht nachweisen und ihren Effekt auch gegenüber formalen Merkmalen absichern. Dabei wiesen die Komplexität des Inhalts sowie die Leitidee einer Aufgabe einen signifikanten Effekt auf, während die kognitive Aktivität keinen substantiellen Einfluss auf die empirische Aufgabenschwierigkeit hatte.

#### 2.3.2.4 *Prozessorientierte Aufgabenmerkmale*

Als prozessorientierte Merkmale werden solche bezeichnet, von denen auf während der Bearbeitung der Aufgaben ablaufende kognitive Prozesse geschlossen wird. Da bei ihrer Zuweisung ein hohes Maß an Interpretation notwendig ist, werden sie als hoch-inferent eingestuft (Kauertz, 2008). Obwohl die Analyse der prozessorientierten Merkmale nützliche Informationen

über die Bearbeitung von Aufgaben liefert, wird sie häufig kritisch betrachtet. Dies liegt daran, dass oft kein eindeutiger Lösungsweg einer Aufgabe identifizierbar ist, da bestimmte Aufgabenmerkmale abhängig vom jeweiligen Bearbeiter unterschiedliche Prozesse auslösen können. Zudem erweist sich die empirische Zuordnung von Aufgaben zu Prozessen häufig als schwierig, und besonders die Analyse genereller kognitiver Prozesse ist für die fachdidaktische Forschung von geringer Relevanz.

### **2.3.3 Merkmale des Kontexts**

#### *2.3.3.1 Kontextmerkmale und Erfassung von Schülervorstellungen*

Aufgaben bestehen im Regelfall aus zwei Komponenten (Rost, 1996): Der Item- bzw. Aufgabenstamm gibt eine Situation vor, in der die Person das von ihr geforderte Testverhalten zeigt. Die zweite Komponente einer Aufgabe ist das Antwortformat und dient zur Registrierung des Testverhaltens. Beide Komponenten wirken auf die Schwierigkeit einer Aufgabe. Der Itemstamm, verstanden als der Kontext, vor dessen Hintergrund eine Aufgabe bearbeitet wird, beeinflusst das Antwortverhalten.

In Instrumenten zur Erfassung naturwissenschaftlicher Kompetenz wird im Itemstamm häufig ein konkretes naturwissenschaftliches Phänomen beschrieben, dessen Erklärung von den Probanden gefordert wird (vgl. z.B. Hardy et al., 2010; Viering et al., 2010). Diese Form der Erfassung ist begründet in der kontext- bzw. situationsbezogenen Definition von Kompetenz (Klieme & Leutner, 2006; Weinert, 2002; siehe Kapitel 2.1.1). Gerade allerdings, wenn es um die Erhebung naturwissenschaftlicher Vorstellungen von Schülern geht, ist bereits bekannt, dass diese vom Kontext beeinflusst werden (siehe Kapitel 2.2.3). Daher scheint es sinnvoll, eine systematische Untersuchung von Kontextmerkmalen und ihrer Wirkung auf die Erfassung naturwissenschaftlicher Schülervorstellungen anzustreben (siehe auch Kapitel 2.2.6).

Entsprechende Merkmale können in der Oberflächenstruktur von Situationen zu finden sein und sind somit niedrig-inferent, d.h. mit einem hohen Maß an Objektivität, aus den Aufgaben abzuleiten. Im Hinblick auf die Untersuchungen zur Kontextualität von Schülervorstellungen kommen dabei v.a. solche Merkmale in Frage, denen empirisch bereits ein Zusammenhang mit Schülerkonzepten nachgewiesen wurde (siehe Kapitel 2.2.3, 2.2.4.2 und 2.2.5.2). Die Kategorisierung der Kontexte anhand dieser Merkmale kann dabei von der fachlich-normativen Einordnung abweichen (diSessa, 2008), d.h. die Kontextmerkmale unterscheiden ggf. zwischen Aufgaben, in denen aus fachlicher Sicht dasselbe Phänomen beschrieben wird.

Die Untersuchung entsprechender Merkmale des Kontexts könnte zur Aufklärung der Schwierigkeitsvarianz von Aufgaben zur Erfassung naturwissenschaftlicher Konzepte beitragen. Darüber hinaus könnte sie hilfreiche Hinweise für die Neukonstruktion von Aufgaben und zur Beschreibung von Skalen liefern und Möglichkeiten für den Einsatz von Aufgaben zur Differenzierung im Unterricht aufzeigen (einfache vs. schwierige Aufgaben zur Bearbeitung

desselben Inhalts). Desweiteren bietet die Untersuchung der Kontextmerkmale eine Möglichkeit zwischen den unterschiedlichen Theorien zur Charakterisierung des initialen naturwissenschaftlichen Wissens zu unterscheiden (siehe Kapitel 2.2.2 und 2.2.3).

Da die vorliegende Arbeit sich mit Schülervorstellungen in den Themenbereichen Verdunstung/Kondensation und Schwimmen/Sinken beschäftigt, wird in den nächsten beiden Abschnitten dargestellt, welche Kontextmerkmale in diesen beiden Bereichen potentiell relevant sind. Dazu wird auf empirische Ergebnisse zurück gegriffen, die bereits in Kapitel 2.2.4.2 und 2.2.5.2 dargestellt wurden. Hier soll allerdings gezielt untersucht werden, welche Bedeutung die Ergebnisse für die Umsetzung in Aufgabenmerkmale haben und welche Effekte dieser Merkmale potentiell erwartet werden können.

### *2.3.3.2 Relevante Kontextmerkmale bei Verdunstung und Kondensation*

Die Verdunstung, d.h. der Aggregatzustandswechsel von flüssig zu gasförmig unterhalb des Siedepunktes, ist für Schüler der Grundschule schwieriger zu verstehen als Verdampfung oberhalb des Siedepunktes, z.B. beim Kochen von Wasser (z.B. Bar & Galili, 1994; Bar & Travis, 1991; Johnson, 1998; Tytler, 2000). Zudem nehmen einige Schüler an, dass Verdampfung nur oberhalb des Siedepunktes bzw. bei explizierter Wärmezufuhr stattfinden kann (Costu & Ayas, 2005). Die Temperatur bzw. die Beschreibung einer Wärmequelle in einer Aufgabe könnte also ein Merkmal sein, welches die Schwierigkeit dieser Aufgabe beeinflusst. Besonders betrachtet werden sollten von den Aufgaben mit Wärmequelle noch einmal jene, in denen die Sonne explizit als Wärmequelle angegeben ist, da bekannt ist, dass die Sonne vielmals in Schülervorstellungen eine wichtige Rolle spielt (z.B. Russell et al., 1989). Aufgrund der o.g. Befunde und der fehlerhaften Schülervorstellungen, die mit Verdunstung unterhalb des Siedepunktes verbunden sind, ist anzunehmen, dass Aufgaben, in denen Verdunstung oberhalb des Siedepunktes erläutert werden soll, leichter zu lösen sind als solche unterhalb des Siedepunktes.

Andere Erklärungen für Verdunstungsphänomene, die häufig von Schülern gegeben werden, beziehen sich auf die Sonne als einen möglichen aktiven Agenten bei der Relokalisation von Wasser (Russell et al., 1989) oder darauf, dass Wasser in die Wolken gelangt (Osborne & Cosgrove, 1983; Russell et al., 1989; Tytler, 2000). Diese beiden Aspekte, Sonne und Wolken, können nur in Situationen außerhalb geschlossener Räumlichkeiten zur Erklärung heran gezogen werden, da sie innen nicht vorhanden sind. Tytler (2000) berichtet außerdem davon, dass bei Phänomenen, die draußen situiert waren, mehr Erklärungen geäußert wurden, die sich auf den gesamten Wasserkreislauf beziehen. Die physikalische Örtlichkeit, an der eine Aufgabe zu Verdunstung bzw. Verdampfung situiert ist, könnte also die Schwierigkeit solcher Aufgaben beeinflussen. Da der Wasserkreislauf neben Verdampfungs- auch Kondensations-Phänomene umfasst, ist dieses Merkmal vermutlich auch für Kondensationsaufgaben relevant. Da mit den draußen situierten Phänomenen in den bisherigen Studien häufig fehlerhafte bzw. naive



Vorstellungen assoziiert wurden, ist zu vermuten, dass diese schwieriger korrekt zu erklären sind als drinnen situierte Phänomene.

V.a. bei Aufgaben, in denen das Trocknen von Wäsche erklärt werden soll, war eine häufig geäußerte Vorstellung, das Wasser ziehe in die Wäsche ein, während diese Vorstellung in Bezug auf andere Aufgabenkontexte seltener geäußert wurde (Bar & Galili, 1994; Osborne & Cosgrove, 1983; Russell et al., 1989). Dieser Unterschied ist möglicherweise dadurch zu erklären, dass im Falle trocknender Wäsche kein offensichtlicher Wasserspiegel, im Gegensatz zu Situationen, in denen z.B. ein fallender Pegel in einem Behältnis beschrieben wird, vorhanden ist, sondern das Wasser bereits von dem Stoff aufgesaugt ist. Das Vorhandensein bzw. eben Nicht-Vorhandensein eines Wasserspiegels, im Sinne einer sichtbaren, geschlossenen Wasseroberfläche, könnte somit die Aufgabenschwierigkeit so beeinflussen, dass aufgrund der Unmöglichkeit der Anwendung der Fehlvorstellung des Einziehens Aufgaben ohne sichtbare, geschlossene Wasserfläche schwieriger zu lösen sind.

Bei der Erklärung von Kondensation führen Schüler häufig an, dass Kälte sich in Wasser verwandelt (Driver et al., 1994; Osborne & Cosgrove, 1983; Tytler et al., 2006) oder dass der Gegenstand schwitzt (Osborne & Cosgrove, 1983). In beiden Fällen sind diese Schülervorstellungen mit einer bestimmten Temperatur verbunden: Kälte kann sich nur dann in Wasser verwandeln, wenn irgendetwas tatsächlich kalt ist, z.B. durch Eiswürfel oder kalte Flüssigkeit. Und Wärme ist eine Voraussetzung des Schwitzens. Somit könnte auch im Falle der Kondensation die Temperatur ein Kontextmerkmal sein, welches die Schwierigkeit von Aufgaben beeinflusst. Da in diesem Fall mit hoher und niedriger Temperatur eine gängige Fehlvorstellung assoziiert ist, kann keine Wirkrichtung dieses Merkmals auf die Aufgabenschwierigkeit abgeleitet werden.

Da v.a. jüngere Schüler noch kein Verständnis von Verdunstung bzw. des gasförmigen Zustandes haben (Bar & Travis, 1991), fällt es ihnen schwer, die Rolle der umgebenden Luft bzw. des darin gelösten gasförmigen Wassers für die Kondensation zu verstehen. Daher könnte das Vorhandensein einer potentiellen offensichtlichen Wasserquelle die Bearbeitung von Aufgaben beeinflussen. Auch die Vorstellung, dass Wasser durch Materie hindurch, z.B. durch eine Glaswand, nach außen gelange und dort Tröpfchen bilde, ist nur dann plausibel, wenn sich im Innern des fraglichen Gefäßes tatsächlich eine Flüssigkeit, also eine mögliche Quelle des kondensierten Wassers, befindet. Da ohne eine potentielle Quelle des Wassers zur korrekten Erklärung auf das in der Luft befindliche Wasser zurück gegriffen werden muss und ein Verständnis davon erst nach und nach von Grundschulern erlangt werden kann (siehe Kapitel 2.2.4.1), sind Aufgaben ohne sichtbare Wasserquelle vermutlich schwieriger zu lösen als solche mit.

Typische Situationen, in denen Kondensation im Alltag beobachtet werden kann, sind z.B. Kondensation am Spiegel beim Duschen, am Fenster während eines kalten Wintertages oder an Gefäßen, die kalte Getränke enthalten (vgl. z.B. Johnson, 1998; Osborne & Cosgrove, 1983; Tytler,

2000). Andere Kondensations-Situationen beziehen sich auf den sichtbaren „Dampf“, beispielsweise beim Kochen oder bei Nebel. Diese beiden Gruppen von Situationen unterscheiden sich deutlich dadurch, dass sich bei der ersten Gruppe Tröpfchen auf einer Fläche bilden, während im zweiten Fall keine Fläche zu sehen ist, sondern der Wasserdampf, also gasförmiges Wasser, an Kondensationskernen in der Luft zu flüssigem Wasser kondensiert. Gerade die Interpretation dieses alltagssprachlichen „Dampfes“ fällt Schülern häufig schwer; sie verwechseln ihn mit Luft oder Wärme (Bar & Travis, 1991). Diese Unterscheidung von Situationen könnte ebenfalls einen entscheidenden Einfluss auf die Aufgabenschwierigkeit haben. Wie beim vorgehenden Merkmal ist auch bei diesem Merkmal zu erwarten, dass der Aufgabentyp, der ein Verständnis davon erfordert, dass das gasförmige Wasser in der Luft ist, also der Typ ohne sichtbare Kondensationskerne, schwieriger zu lösen ist.

#### *2.3.3.3 Relevante Kontextmerkmale bei Schwimmen und Sinken*

Bei der Erklärung von Schwimmen bzw. Sinken werden von Schülern häufig unterschiedliche Gründe genannt, die sich auf bestimmte Attribute einzelner Objekte beziehen (Biddulph & Osborne, 1984; Engelen et al., 2002; Pollmeier et al., 2011; Tytler & Peterson, 2004). Die genannten beachteten Charakteristika sind dabei meist äußerlich wahrnehmbar (Hsin & Wu, 2011). Häufig beziehen sich die genannten Erklärungen für das Schwimmverhalten auf das Gewicht von Objekten (Biddulph & Osborne, 1984; Butts et al., 1993; Tytler & Peterson, 2004). Schüler nehmen z.B. an, dass nur leichte Dinge schwimmen und schwere untergehen, ohne dabei das Gewicht an der Größe des Gegenstandes zu relativieren. Da das Schwimmverhalten allerdings nicht allein vom Gewicht bestimmt wird, sondern auch schwere Objekte, wie z.B. ein Schiff, schwimmen, könnten entsprechende Aufgabenstellungen für Schüler, die erwarten, dass schwere Dinge immer untergehen, überraschend oder unplausibel sein und so die Schwierigkeit einer Aufgaben beeinflussen. Dabei sind Aufgaben, in denen es um besonders schwere Gegenstände geht, vermutlich schwieriger zu lösen, da sie konträr zu dieser gängigen Fehlvorstellung sind.

Viele Schüler, die nicht auf das Gewicht eines Gegenstandes fokussieren, beachten die Größe (Butts et al., 1993; Hsin & Wu, 2011; Pollmeier et al., 2011). Da auch die Größe allein nicht für eine zuverlässige Vorhersage des Schwimmverhaltens geeignet ist, könnte auch eine Variation dieses Aufgabenmerkmals die Schwierigkeit beeinflussen. Entsprechend des vorherigen Merkmals ist auch in diesem Fall anzunehmen, dass Aufgaben, in denen besonders große Gegenstände thematisiert werden, besonders schwer zu lösen sind, da sie im Widerspruch zur gängigen Fehlvorstellung stehen, dass kleine Dinge schwimmen können.

Neben Attributen des Gegenstands selbst fokussieren Schüler in ihren Erklärungen des Schwimmverhaltens bei Hohlkörpern häufig auf die im Gegenstand befindliche Luft (Engelen et al., 2002; Fischer, 2008; Havu-Nuutinen, 2005). Zumeist führt ihre Erklärung damit schon in eine richtige Richtung, da die Größe des Hohlraums eines Gegenstandes die relative Dichte eines

Gegenstandes im Vergleich zum Wasser verringert und in einigen Fällen, z.B. bei Schiffen aus Metall, das Schwimmen ermöglicht. Dennoch geraten die Schüler auch mit dieser Erklärung noch an Grenzen, da es auch hohle Gegenstände gibt, die untergehen. Darüber hinaus ist dieser Erklärungsansatz auf Vollkörper nicht direkt übertragbar. Ein abweichendes Antwortverhalten und damit eine veränderte Aufgabenschwierigkeit sind auch bei unterschiedlichen Ausprägungen dieses Merkmals (Hohlkörper vs. Vollkörper) denkbar. Bei Vollkörpern gibt es keine im Gegenstand eingeschlossene Luft, die potentiell Assoziationen in Richtung einer korrekten Erklärung (Verringerung der Dichte) evozierten kann. Daher sind entsprechende Aufgaben vermutlich schwieriger zu lösen als solche mit Hohlkörpern.

Bei der Erfassung des Dichteverständnisses zeigte sich, dass das konkrete Verhältnis der Dichte eines Gegenstandspaares zueinander das Antwortverhalten in Interviews beeinflusste (Fassouloupoulos et al., 2003; Smith et al., 1985). Dabei waren speziell Aufgaben mit inkongruentem Dichteverhältnis schwierig zu lösen, d.h. Aufgaben, in denen der absolut leichtere Gegenstand die höhere Dichte aufwies. In Aufgaben zu Dichte könnte die Kongruenz bzw. Inkongruenz von Gegenstandspaares also ein schwierigkeiterzeugendes Merkmal sein.

In Bezug auf die Verdrängung haben Schüler häufig noch nicht verstanden, dass die Verdrängung einer Flüssigkeit durch ein Objekt von dessen Volumen und nicht dessen Gewicht abhängt, sofern das Objekt sinkt (z.B. Pollmeier et al., 2011). Da die Schüler allerdings in der Regel nach den Gründen für das Ansteigen bzw. Überlaufen des Wassers beim Eintauchen eines einzelnen Objektes gefragt werden, ist vorstellbar, dass das Antwortverhalten anders ausfällt, wenn nach dem Vergleich des Wasseranstiegs zweier unterschiedlicher Objekte gefragt wird, was somit ein schwierigkeiterzeugendes Merkmal darstellen könnte. Über die Wirkrichtung dieses Merkmals können keine Vermutungen angestellt werden.

### **3 Zielsetzung, Fragestellungen und Hypothesen**

Wie in Kapitel 2.1.2.5 ausgeführt wurde, ist die Modellierung naturwissenschaftlicher Kompetenz in der Grundschule erst angelaufen. Als Ansatzpunkt für die Formulierung plausibler Kompetenzmodelle für diesen Bereich können die Erkenntnisse der Forschung zu Konzeptwechsel und Schülervorstellungen dienen (siehe Kapitel 2.2.1). Diese liefern wichtige Informationen über die Vorstellungen von Schülern zu naturwissenschaftlichen Themen vor dem Unterricht und die Veränderung dieser Vorstellungen hin zu angemesseneren Vorstellungen. Bereits jetzt wird das Paradigma des Conceptual Change von der Fachdidaktik der Naturwissenschaften genutzt. Der Rückgriff auf diese bereits vorliegenden Forschungsergebnisse könnte in Bezug auf die Kompetenzmodellierung insbesondere zu einer höheren Validität der Kompetenzmodelle sowie zugehöriger Testinstrumente beitragen.

Eine besondere Eigenschaft von naturwissenschaftlichen Schülervorstellungen, die im Rahmen der Konzeptwechselforschung bereits häufig nachgewiesen wurde, ist die kontextualisierte Natur dieser Vorstellungen: Häufig werden inkonsistente Vorstellungen von Schülern in Abhängigkeit von der Variation von Kontextmerkmalen geäußert (siehe Kapitel 2.2.3). Diese Kontextabhängigkeit wird allerdings in aktuellen Ansätzen zur Modellierung naturwissenschaftlicher Kompetenz von Grundschulern nicht ausreichend berücksichtigt (siehe Kapitel 2.1.2.5). Dies ist besonders verwunderlich, da Kompetenz bezogen auf spezifische Situationen bzw. Kontexte definiert wird (siehe Kapitel 2.1.1.).

Bei der Erfassung von Kompetenzen können die Aufgabenstämme als die Kontexte verstanden werden, vor deren Hintergrund Schüler ihre Vorstellungen mittels des Testverhaltens kundtun (siehe Kapitel 2.3.3.1). Eine systematische Untersuchung der Merkmale dieser Kontexte und ihres Einflusses auf die Erfassung naturwissenschaftlicher Schülervorstellungen kann zur Erklärung der Schwierigkeitsvarianz von Aufgaben beitragen, hilfreiche Hinweise für die Neukonstruktion von Aufgaben und zur Beschreibung von Skalen liefern sowie Möglichkeiten für den Einsatz unterschiedlich schwieriger Aufgaben zur Differenzierung im Unterricht aufzeigen.

Auch für die in dieser Arbeit betrachteten Themenbereiche Verdunstung/Kondensation und Schwimmen/Sinken wurden aufgaben- bzw. kontextabhängige Schülerantworten berichtet (z.B. Biddulph & Osborne, 1984; Hsin & Wu, 2011; Russell et al., 1989; Tytler, 2000). Die meisten Autoren nutzten dafür Interviewstudien (siehe Kapitel 2.2.4.2 und 2.2.5.2), in denen unterschiedliche Aufgabenstellungen verwendet wurden. Allerdings war die Untersuchung der Kontextabhängigkeit des Antwortverhaltens nicht immer intendiert und es wurde nicht klar definiert, welche Kriterien bedeutsam für die Unterscheidung von Situationen waren.

Die Zielsetzung meiner Arbeit ist daher die Identifikation von Kontextmerkmalen in schriftlichen, größtenteils geschlossenen Aufgaben zur Erfassung naturwissenschaftlicher Kompetenz von Grundschulern in den o.g. Themenbereichen und die Analyse von deren Einfluss auf die

Aufgabenschwierigkeit. Meine Intention ist damit einen Beitrag zur Aufklärung und Verbesserung der Validität eines entsprechenden Testinstruments zu leisten. Die sich daraus ergebenden Fragestellungen und Hypothesen, die für meine Untersuchungen leitend waren, werden im Folgenden beschrieben.

In den Kapiteln 2.3.3.2 und 2.3.3.3 wurde heraus gearbeitet, welche Kontextmerkmale für die beiden untersuchten Themenbereiche bzw. die jeweils relevanten Konzepte potentiell ein verändertes Antwortverhalten und damit eine unterschiedliche Aufgabenschwierigkeit bewirken könnten. Für Aufgaben zur Verdunstung bzw. Verdampfung sind dies:

- das Vorhandensein einer Wärmequelle unter besonderer Berücksichtigung von Aufgaben, in denen die Sonne diese Quelle darstellt,
- der Ort, an dem die Aufgabe situiert ist, wobei zwischen Aufgaben innerhalb und außerhalb geschlossener Räumlichkeiten differenziert wird, und
- die Sichtbarkeit einer geschlossenen Wasseroberfläche bzw. eines Wasserspiegels.

Für Kondensationsphänomene wurden folgende Merkmale als möglicherweise bedeutsam heraus gearbeitet:

- der Ort, wobei wie für den Bereich Verdunstung zwischen drinnen und draußen unterschieden wird,
- die Temperatur, die in den Situationen entweder erhöht (z.B. bei kochendem Wasser) oder abgesenkt (z.B. bei einem Gefäß, das mit einem kalten Getränk gefüllt ist) sein kann,
- das Vorhandensein einer möglichen Quelle für entstehende Wassertröpfchen und
- die Unterscheidung von Aufgaben, in denen es um die Erklärung von Kondensation an Flächen geht und solchen, in denen Wassertröpfchen als sichtbarer „Dampf“ in der Luft kondensieren.

Bei dem für die Erklärung des Schwimmverhaltens relevanten Verständnis der Auftriebskraft könnten folgende Merkmale das Antwortverhalten beeinflussen:

- das Gewicht eines Gegenstandes, dessen Schwimmverhalten erklärt werden soll,
- die Größe des Gegenstandes und
- die Unterscheidung zwischen Hohl- und Vollkörpern.

Für Aufgaben, die das Verständnis der Dichte erfassen, soll die Kongruenz der Aufgaben als Kontextmerkmal untersucht werden. Hierbei wird zwischen kongruenten Aufgaben, in denen der schwerere Gegenstand die höhere Dichte hat, und inkongruenten Aufgaben, in denen der leichtere Gegenstand die höhere Dichte hat, unterschieden.

Ob ein Vergleich zwischen zwei Gegenständen vorgenommen werden soll oder die Verdrängung in einer Flüssigkeit durch einen einzelnen Gegenstand ohne Vergleich zu einem weiteren erklärt werden soll, wird als Kontextmerkmal in den Aufgaben zu Verdrängung untersucht.

Hinsichtlich dieser Kontextmerkmale lautet die erste Frage, der in dieser Arbeit nachgegangen werden soll:

**Frage 1 (F1):**

In welchem Umfang sind unterschiedliche Ausprägungen der verschiedenen Kontextmerkmale in Aufgaben zur Erfassung naturwissenschaftlicher Kompetenz von Grundschulern in den beiden Themenbereichen Verdunstung/Kondensation und Schwimmen/Sinken enthalten?

Für diese erste Frage werden keine gerichteten Hypothesen formuliert. Sie soll eine gezielte Analyse und Beschreibung des eingesetzten Aufgabenmaterials hinsichtlich der untersuchten Merkmale anstoßen.

Für die Frage nach der möglichen Relevanz der Kontextmerkmale für die Modellierung naturwissenschaftlicher Kompetenz von Grundschulern ist die Beantwortung der zweiten Forschungsfrage von Bedeutung:

**Frage 2 (F2):**

Beeinflussen die in den Aufgaben zur Erfassung naturwissenschaftlicher Kompetenz von Grundschulern enthaltenen Kontextmerkmale das Antwortverhalten der Probanden und somit die Aufgabenschwierigkeit?

Für diese zweite Frage werden folgende Hypothesen formuliert:

**Frage 2 – Hypothese 1 (F2-H1):** Die untersuchten Kontextmerkmale zeigen einen bedeutsamen Einfluss auf die Aufgabenschwierigkeit.

**Frage 2 – Hypothese 2 (F2-H2):** Die Richtung des Effekts der Kontextmerkmale entspricht der in Tabelle 1 explizierten angenommenen Wirkrichtung (zur Begründung der angenommenen Wirkrichtungen siehe Kapitel 2.3.3.2 und 2.3.3.3).

**Tabelle 1 Untersuchte Aufgabenmerkmale und deren angenommen Wirkrichtung auf die Aufgabenschwierigkeit**

Be- reich <sup>a</sup>	Aufgaben- merkmal	Angenommene Wirkrichtung		Kodierung <sup>b</sup>
		leicht	schwer	
Formale Merkmale				
alle	Textlänge	kurz < lang		Anzahl der Wörter drei- bis fünffach gestuft dichotom
	Antwortformat	forced choice < multiple select < multiple choice < offen <sup>c</sup>		
	Demonstration	ja < nein		
Kontextmerkmale				
Vd	Wasserspiegel	sichtbar < unsichtbar		dichotom
	Ort	drinnen < draußen		dichotom
	Temperatur	warm/heiß < Sonne < kalt		zwei- bzw. dreifach gestuft

K	Kondensationskern Temperatur Quelle Ort	Fläche < Dampf warm ? <sup>d</sup> kalt sichtbar < unsichtbar drinnen < draußen	dichotom dichotom dichotom dichotom
A	Größe Gewicht Hohl	klein < groß leicht < schwer hohl < gefüllt	dichotom dichotom dichotom
D	Kongruenz	kongruent < eine Dimension (Gewicht, Dichte) konstant gehalten < inkongruent	dreifach gestuft

V	Vergleich	Einzelurteil ? Vergleich	dichotom
<i>Anmerkungen.</i> <sup>a</sup> Vd=Verdunstung, K=Kondensation, A=Auftrieb, D=Dichte, V=Verdrängung; <sup>b</sup> Hier wird bereits auf die Kodierung der Merkmale in den anschließend dargestellten Studien vorgegriffen. Diese wird im jeweiligen Methodenteil noch ausführlich dargestellt. <sup>c</sup> Nicht jedes Format kam bei allen untersuchten Komponenten vor, z.B. gab es keine offenen Aufgaben für den Themenbereich Verdunstung und Kondensation. Eine Aufgabe, bei der eine graphische Darstellung als Antwort verlangt wurde, wurde unter den offenen Aufgaben subsummiert. <sup>d</sup> Bei diesen Aufgabenmerkmalen konnte nicht plausibel auf eine mögliche Richtung des Effekts auf die Aufgabenschwierigkeit geschlossen werden.			

Neben einem allgemeinen Effekt wurde ein abnehmender Einfluss der Kontextmerkmale bei steigender wissenschaftlicher Angemessenheit der Schülervorstellungen nachgewiesen (vgl. Engel Clough & Driver, 1986; siehe Kapitel 2.2.3). Auch der Ansatz der Fragmentierung von Schülervorstellungen (z.B. diSessa, 2008) geht von einer stärkeren Vernetzung und Systematisierung des Wissens bei zunehmender Expertise der Lernenden aus. Eine mögliche Interaktion zwischen Kontextmerkmalen und der wissenschaftlichen Angemessenheit einer Erklärung könnte sich im Falle des Projektes Science-P, in dessen Rahmen die vorliegende Arbeit eingebettet ist, in Bezug auf die angenommenen Kompetenzniveaus des postulierten Modells ausdrücken (vgl. Kapitel 2.1.2.4). Folgende Fragestellung bezieht sich daher auf mögliche Interaktionen zwischen Kontextmerkmalen und dem Niveau einer Erklärung:

### Frage 3 (F3):

Verändert sich der Einfluss der Kontextmerkmale auf die Aufgabenschwierigkeit in Abhängigkeit von der wissenschaftlichen Angemessenheit bzw. dem Niveau einer Aufgabe?

In Übereinstimmung mit Befunden, die abnehmende Kontextualisierung mit zunehmendem Antwortniveau belegen (z.B. Engel Clough & Driver, 1986), wird zu dieser Frage folgende Hypothese aufgestellt:

**Frage 3 – Hypothese 1 (F3-H1):** Die Stärke des Effekts eines Kontextmerkmals auf die Aufgabenschwierigkeit verringert sich mit zunehmendem Antwortniveau, d.h. es gibt eine signifikante Interaktion zwischen Kontextmerkmal und Niveau.

Die dieser Arbeit zugrunde gelegten Daten wurden mit Aufgaben aus dem Projekt Science-P (siehe Kapitel 2.1.2.4) erhoben. Bei der Konstruktion dieser Aufgaben wurden unterschiedliche Aufgabenformate realisiert. Weitere formale Aufgabenmerkmale wurden nicht systematisch bei der

Aufgabenentwicklung berücksichtigt. Es ist allerdings bekannt, dass diese formalen, konstruktirrelevanten Merkmale einen Einfluss auf die Aufgabenschwierigkeit haben können. Daher sollen sie im Sinne von Kontrollvariablen bei der Untersuchung des Einflusses von Aufgabenmerkmalen mit folgender Forschungsfrage berücksichtigt werden:

**Frage 4 (F4):**

Welche Bedeutung haben formale Aufgabenmerkmale in Aufgaben zur Erfassung naturwissenschaftlicher Kompetenz von Grundschulern im Vergleich zu Kontextmerkmalen?

Formale Aufgabenmerkmale, wie das Antwortformat oder die Textlänge, haben oft einen großen Anteil an der Aufgabenschwierigkeit und da besonders in der Grundschule noch nicht alle Voraussetzungen für einen kompetenten Umgang mit den formalen Anforderungen von Aufgaben gegeben sind, z.B. weil das Leseverständnis noch nicht voll entwickelt ist (Carr, Kurtz, Schneider, Turner & Borkowski, 1989), werden in Zusammenhang mit Frage 4 folgende Hypothesen aufgestellt:

**Frage 4 – Hypothese 1 (F4-H1):** Die untersuchten formalen Aufgabenmerkmale zeigen einen bedeutsamen Einfluss auf die Aufgabenschwierigkeit.

**Frage 4 – Hypothese 2 (F4-H2):** Die Richtung des Effekts der formalen Merkmale entspricht der in Tabelle 1 explizierten angenommenen Wirkrichtung.

Darüber hinaus können die formalen Aufgabenmerkmale als Kontrollvariablen verstanden werden und der Ausmaß ihres Einflusses mit dem der Kontextmerkmale verglichen werden. Da die formalen Merkmale allerdings keinerlei inhaltlichen Bezug zum untersuchten Konstrukt aufweisen, die Kontextmerkmale aber sowohl für die Modellierung von Kompetenzen wie auch für Schülervorstellungen eine nachgewiesene Bedeutung haben, wird folgende letzte Hypothese formuliert:

**Frage 4 – Hypothese 3 (F4-H3):** Die Stärke des Effekts der formalen Merkmale auf die Aufgabenschwierigkeit übersteigt nicht die Stärke des Effekts der Kontextmerkmale, d.h. die Varianzaufklärung an der Aufgabenschwierigkeit durch Kontextmerkmale ist größer.

In den folgenden Abschnitten werden zwei empirische Untersuchungen zur Beantwortung der Fragestellungen beschrieben. Die Ergebnisse der Studien werden vorgestellt und hinsichtlich der Fragen und Hypothesen diskutiert.

Bei Studie 1 handelt es sich um eine Querschnittstudie die im Rahmen des Projekts Science-P zur Überprüfung des postulierten Kompetenzmodells durchgeführt wurde. Die in Frage stehenden Merkmale der Aufgabenkontexte wurden in diesem Aufgabenpool nicht systematisch variiert.

Für Studie 2 wurden die Aufgaben aus Studie 1 hinsichtlich der relevanten Merkmale überarbeitet. Dabei wurden die Kontextmerkmale der Aufgaben systematisch variiert, während andere, v.a. formale, Merkmale konstant gehalten wurden.



## **4 Studie 1: Querschnittstudie**

### **4.1 Einbettung der Studie 1 in das Projekt Science-P**

In dem Projekt Science-P wurden, ausgehend vom postulierten Kompetenzmodell (siehe Kapitel 2.1.2.4), Aufgaben zur Erfassung der Kompetenz in den Bereichen *Naturwissenschaftliches Wissen* und *Wissen über Naturwissenschaften* in der Grundschule entwickelt und optimiert. Zur Validierung des Kompetenzmodells wurde nach Abschluss der Instrumententwicklung eine Querschnittstudie in der zweiten, dritten und vierten Jahrgangsstufe der Grundschule durchgeführt. Eine Überprüfung weiterer Merkmale von Aufgaben, die ohne direkten Bezug zum postulierten Kompetenzmodell sind, war in dem Projekt nicht angestrebt. Für die vorliegende Arbeit wurde auf die Daten aus dieser Querschnittstudie zur Dimension *Naturwissenschaftliches Wissen* zurückgegriffen. Die genaue Entwicklung und Auswahl der Testitems wird in Kapitel 4.3.1.1 beschrieben.

### **4.2 Präzisierung und Einschränkung der Fragestellungen und Hypothesen für Studie 1**

In dieser ersten Studie wurde auf Aufgaben zurück gegriffen, die nicht speziell für die Untersuchung der in Kapitel 3 explizierten Fragen entwickelt wurden. Dennoch war die Untersuchung aller in Kapitel 3 aufgeworfenen Fragestellungen möglich und wurde mit dieser Studie angestrebt. Eine notwendige Einschränkung ergab sich dabei für die Frage 3, die sich auf die Untersuchung von Interaktionen bezieht. Für die Analyse der Daten dieser Fragestellung ist es nötig, dass alle Zellen des entstehenden Designs von Aufgaben besetzt sind. Beispielsweise sollten für das Aufgabenmerkmal Ort für alle möglichen Ausprägungen dieses Merkmals (drinnen, draußen) Aufgaben auf jedem Kompetenzniveau vorliegen. Die Aufgabenanalyse hat gezeigt, dass das nicht für alle untersuchten Merkmale der Fall ist (siehe Kapitel 4.4.1). Die Untersuchung dieser Frage wurde deshalb auf die Merkmale eingeschränkt, bei denen diese Voraussetzung gegeben ist. In Studie 1 wurden also folgende Fragestellungen und Hypothesen untersucht:

#### **Frage 1 (F1):**

In welchem Umfang sind unterschiedliche Ausprägungen der verschiedenen Kontextmerkmale in Aufgaben der Querschnittstudie zur Erfassung naturwissenschaftlicher Kompetenz von Grundschülern in den beiden Themenbereichen Verdunstung/Kondensation und Schwimmen/Sinken enthalten?

#### **Frage 2 (F2):**

Beeinflussen die in den Aufgaben zur Erfassung naturwissenschaftlicher Kompetenz von Grundschülern enthaltenen Kontextmerkmale das

Antwortverhalten der Probanden in der Querschnittstudie und somit die Aufgabenschwierigkeit?

**Frage 2 – Hypothese 1 (F2-H1):** Die untersuchten Kontextmerkmale zeigen einen bedeutsamen Einfluss auf die Aufgabenschwierigkeit in der Querschnittstudie.

**Frage 2 – Hypothese 2 (F2-H2):** Die Richtung des Effekts der Kontextmerkmale in der Querschnittstudie entspricht der in Tabelle 1 (siehe Kapitel 3) explizierten angenommenen Wirkrichtung.

**Frage 3 (F3):**

Verändert sich der Einfluss der Kontextmerkmale auf die Aufgabenschwierigkeit in Abhängigkeit von der wissenschaftlichen Angemessenheit bzw. dem Niveau einer Aufgabe?

**Frage 3 – Hypothese 1 (F3-H1):** Die Stärke des Effekts eines Kontextmerkmals auf die Aufgabenschwierigkeit verringert sich mit zunehmendem Antwortniveau, d.h. es gibt in der Querschnittstudie signifikante Interaktionen zwischen Kontextmerkmalen und Niveaus bei Merkmalen, bei denen alle möglichen Kombinationen von Merkmal und Niveau vorliegen.

**Frage 4 (F4):**

Welche Bedeutung haben formale Aufgabenmerkmale in den Aufgaben der Querschnittstudie zur Erfassung naturwissenschaftlicher Kompetenz von Grundschulern im Vergleich zu Kontextmerkmalen?

**Frage 4 – Hypothese 1 (F4-H1):** Die untersuchten formalen Aufgabenmerkmale zeigen einen bedeutsamen Einfluss auf die Aufgabenschwierigkeit in der Querschnittstudie.

**Frage 4 – Hypothese 2 (F4-H2):** Die Richtung des Effekts der formalen Merkmale entspricht in der Querschnittstudie der in Tabelle 1 (siehe Kapitel 3) explizierten angenommenen Wirkrichtung.

**Frage 4 – Hypothese 3 (F4-H3):** Die Stärke des Effekts der formalen Merkmale auf die Aufgabenschwierigkeit übersteigt in der Querschnittstudie nicht die Stärke des Effekts der Kontextmerkmale, d.h. die Varianzaufklärung an der Aufgabenschwierigkeit durch Kontextmerkmale ist größer.

### 4.3 Methode Studie 1

#### 4.3.1 Anlage und Aufbau der Untersuchung in Studie 1

##### 4.3.1.1 Instrumente zur Erfassung des Naturwissenschaftlichen Wissens in Studie 1

Für die vorliegende Arbeit wurde das Wissen der Schüler in den Themenbereichen *Schwimmen/Sinken* und *Verdunstung/Kondensation* mit geschlossenen, schriftlichen Aufgaben erfasst. Diese wurden im Rahmen des Projektes Science-P für die Dimension *Naturwissenschaftliches Wissen* (siehe Kapitel 2.1.2.3) in einem mehrschrittigen Prozess, angelehnt an die von Wilson (2005) beschriebene konstruktbasierte Instrumententwicklung, entwickelt (vgl. Kleickmann et al., 2010). Bei diesem Ansatz bildet den Ausgangspunkt der Entwicklung eine sog. *Construct Map*, in der unterschiedliche Ausprägungen der Kompetenz bei Personen, sowie analog dazu entsprechende Merkmale für Aufgaben, operationalisiert werden. Die auf Grundlage dieser Konstruktbeschreibung entworfenen Aufgaben sollen dann in mehreren Schritten überprüft und die anfängliche Beschreibung auf Grundlage der gewonnenen Ergebnisse kritisch evaluiert werden (Wilson, 2005).

In Science-P wurden zur Operationalisierung der zu messenden Kompetenz zunächst die Zellen des postulierten Struktur-Niveau-Modells (siehe Kapitel 2.1.2.4) mit Rückgriff auf Erkenntnisse der Schülerkonzept-Forschung (vgl. für Schwimmen/Sinken z.B. Engelen et al., 2002; Hardy & Stern, 2010; Smith, Maclin, Grosslight & Davis, 1997; für Verdunstung/Kondensation z.B. Bar & Travis, 1991; Tytler et al., 2006) spezifiziert und eine Gewichtung der einzelnen Komponenten am späteren Gesamttest festgelegt. Das Verständnis von Dichte und Verdrängung sollte jeweils einen geringeren Anteil daran haben als das Verständnis von Auftrieb sowie dessen Nutzung zur Erklärung von Schwimmen bzw. Sinken. Verdunstung und Kondensation sollten gleichwertig im Test berücksichtigt werden.

Zu Beginn des Projektes wurde, neben den bereits beschriebenen Komponenten, für den Themenbereich *Verdunstung/Kondensation* außerdem noch der natürliche Wasserkreislauf besonders mit den Aspekten Wolken und Regen untersucht. Aus inhaltlichen und methodischen Gründen wurde im Laufe der Instrumententwicklung von der Projektgruppe entschieden, diese Aspekte wieder fallen zu lassen. Daher sind keine entsprechenden Aufgaben im endgültigen Testinstrument. Ebenso wurde anfänglich versucht auch für die Komponenten Dichte und Verdrängung ein Niveau der *Zwischenvorstellungen* abzubilden. Aber obwohl entsprechende Vorstellungen von Schülern in mündlichen Interviews vereinzelt genannt wurden (vgl. Pollmeier et al., 2011), ist es nicht gelungen, diese trennscharf in den Tests abzubilden. Nach sorgfältiger inhaltlicher Abwägung ist dieses Niveau für die Komponenten Dichte und Verdrängung wieder fallen gelassen worden.

Zu den Antwortformaten der ursprünglich entwickelten 120 schriftlichen Aufgaben gehörten hauptsächlich solche, bei denen die beste aus zwei (*forced choice*, FC) oder mehreren (*multiple*

*choice*, MC) Antwortalternativen gewählt werden sollte und solche bei denen mehrere Antwortalternativen separat als richtig oder falsch beurteilt werden sollten (*multiple select*, MS). Die vorgegebenen Antworten können jeweils einem der postulierten Niveaus zugeordnet werden. Außerdem wurden vereinzelt Aufgaben mit anderem Antwortformat (offen, graphisch) entwickelt. Die meisten Aufgaben erforderten die Erklärung eines alltagsnahen Phänomens (z.B. Wäsche trocknet auf der Leine) oder eines einfachen Versuchs (z.B. zwei Gläser, gefüllt mit einer identischen Wassermenge, werden an unterschiedlichen Orten aufgestellt und die Menge des verbleibenden Wassers nach einer Woche verglichen). In einigen Aufgaben war eine Vorhersage zu leisten (z.B. aus welchem Gefäß in einer Woche am meisten Wasser verdunstet).

Im Anschluss an die Entwicklung wurden die Aufgaben in einer Pre-Pilotierungsstudie mit 42 Klassen der zweiten und vierten Jahrgangsstufe einer ersten empirischen Prüfung unterzogen (Kleickmann et al., 2010). Hierbei wurde vor allem auf die Verständlichkeit der Aufgaben sowie die Durchführbarkeit als Gruppentest geachtet. Die Daten dieser Untersuchung wurden außerdem hinsichtlich psychometrischer Kriterien (Schwierigkeit, Trennschärfe, Boden- und Deckeneffekte, Reliabilität) analysiert. Darüber hinaus waren in dieser Studie erste Hinweise auf die Sensitivität des Instruments gegenüber Unterricht und Kompetenzentwicklung zu finden (Kleickmann et al., 2010).

In einem nächsten Schritt wurde die Übereinstimmung der Schülerantworten auf eine Auswahl von Aufgaben mit den Antworten auf entsprechende mündliche Interviewaufgaben anhand einer Stichprobe von drei Klassen der dritten Jahrgangsstufe geprüft (Pollmeier et al., 2011). Es konnten substantielle Zusammenhänge zwischen den schriftlichen und mündlichen Instrumenten gefunden werden. Dabei gelang es den Schülern im Fragebogen leichter Antworten auf dem mittleren oder höchsten Niveau zu geben als im Interview.

Schließlich wurden die psychometrischen Eigenschaften von 89 selektierten Aufgaben in einer Item-Pilotierungsstudie mit 57 Klassen überprüft. Auf Basis der Ergebnisse wurden die Aufgaben vor der anschließenden Querschnittstudie einer letzten Selektion und Überarbeitung unterzogen.

Bei allen Schritten der Testentwicklung wurden statistische sowie inhaltliche Kriterien zur Entscheidung über die Selektion von Aufgaben herangezogen und darauf geachtet, die a priori beschlossene Gewichtung der inhaltlichen Komponenten beizubehalten, um so eine inhaltliche Veränderung des untersuchten Konstrukts zu vermeiden. Eine intendierte inhaltliche Veränderung der Construct Map resultierte aus der Schwierigkeit der Abbildung des Niveaus der Zwischenvorstellungen für die Komponenten Verdrängung und Dichte. Die Ergebnisse der Studien zeigten einerseits die Schwierigkeit einer trennscharfen Erfassung dieses spezifischen Aspekts der naturwissenschaftlichen Kompetenz. Andererseits führte die altersangemessene Umsetzung in schriftliche Testaufgaben zu unplausiblen Formulierungen. Daher beschloss die Projektgruppe für den weiteren Verlauf auf eine Erhebung dieser Facette zu verzichten. Aus der letzten Vorstudie, der Item-Pilotierungsstudie, wurden neben Erkenntnissen über die Aufgaben außerdem Schlüsse für


die optimale Verteilung der Aufgaben auf das Rotationsdesign der Testhefte gezogen und konnten in der vorliegenden Querschnittstudie umgesetzt werden.

**Abbildung 1** Ankeraufgabe zur Komponente Verdunstung im Antwortformat *multiple select*

**Kaltes Glas**  
*Instruktion des Testleiters:*  
*Du füllst ein Glas mit Leitungswasser und Eiswürfeln. Das Glas ist zunächst außen ganz trocken, aber nach einem kurzen Moment sieht es so aus wie auf dem Bild: Plötzlich sind Wassertropfen außen auf dem Glas.*  
*Damit du dir das gut vorstellen kannst, werden wir den Versuch jetzt selbst ausprobieren (Demonstration mit den entsprechenden Materialien).*

Plötzlich sind Wassertropfen außen auf dem Glas.  
 Wie sind die Wassertropfen außen an das Glas gekommen?

Kreuze nach jeder Antwort ‚Richtig‘ oder ‚Falsch‘ an!



	Richtig	Falsch
Die Wassertropfen sind durch feine Poren im Glas von innen nach außen gekommen. (NV)	<input type="checkbox"/>	<input type="checkbox"/>
Wasser aus der Luft ist durch die Kälte als Wassertropfen sichtbar geworden. (WV)	<input type="checkbox"/>	<input type="checkbox"/>
Die Wassertropfen sind wegen der Kälte aus der Luft entstanden. (ZV)	<input type="checkbox"/>	<input type="checkbox"/>
Das Wasser aus dem Glas ist jetzt außen dran. (NV)	<input type="checkbox"/>	<input type="checkbox"/>

*Anmerkungen.* NV, naive Vorstellung; ZV, Zwischenvorstellung; WV, wissenschaftliche Vorstellung;

Der resultierende Aufgabenpool für diese Studie umfasste für die Dimension *Naturwissenschaftliches Wissen* 66 Aufgaben mit 149 Items<sup>2</sup>, von denen sieben nur in der dritten und vierten, nicht aber in der zweiten Jahrgangsstufe, eingesetzt wurden, da sie sich in den vorherigen Studien als besonders schwierig erwiesen hatten. Die Aufgaben wurden in einem rotierten Design auf sechs Testhefte verteilt (vgl. Anhang A01). Jedes Testheft<sup>3</sup> umfasste dabei 18 (zweite Klasse) bis 22 (dritte und vierte Klasse) Aufgaben, von denen ein Set bestehend aus 10 Aufgaben als Ankeraufgaben in jedem Testheft vertreten war. Als Ankeraufgaben wurden solche gewählt, die - wenn möglich - Antworten auf allen drei postulierten Kompetenzniveaus enthielten bzw. zuließen und in der Item-Pilotierung besonders zufriedenstellende psychometrische Eigenschaften zeigten. Alle fünf inhaltlichen Komponenten (Dichte, Verdrängung, Auftrieb, Verdunstung, Kondensation) waren zweimal in den Ankeraufgaben vertreten, wobei eine der

<sup>2</sup> Da pro Aufgabenstamm teilweise mehr als eine Antwortreaktion von den Probanden verlangt war (z.B. separate Beurteilung mehrerer Antwortalternativen bei MS-Aufgaben), übersteigt die Anzahl an Items die Anzahl der Aufgaben.

<sup>3</sup> Alle Testhefte sind in der Version für den Testleiter in den elektronischen Anhängen zu finden.

Ankeraufgaben jeweils augenscheinlich komplex und eine einfacher war. Bei der Verteilung der restlichen Aufgaben auf die Aufgabenblöcke bzw. Testhefte wurde darauf geachtet, dass in allen Testheften alle Komponenten sowie alle eingesetzten Antwortformate enthalten waren und dass die Gewichtung der Komponenten in den Testheften annähernd gleich war (vgl. Tabelle 2). Eine exemplarische Aufgabe aus den Ankeraufgaben zur Komponenten Verdunstung im Format *multiple select* ist in Abbildung 1 zu sehen.

**Tabelle 2 Übersicht über eingesetzte Aufgaben und Items in der Querschnittstudie**

Konzept	Aufgabenformat							Gesamt
	<i>forced choice</i>		<i>multiple select</i>	<i>multiple choice</i>		offen	andere	
	NV- ZV <sup>a</sup>	ZV- WV		3 <sup>b</sup>	4 <sup>b</sup>			
Verdunstung	5/ 5 <sup>c</sup>	1/ 1	7/ 28	3/ 3	1/ 1	-/ -	-/ -	17/ 38
Kondensation	5/ 5	4/ 4	7/ 26	1/ 1	-/ -	-/ -	-/ -	17/ 36
Auftrieb	3/ 3	2/ 2	11/ 44	-/ -	-/ -	1/ 3	-/ -	17/ 52
Dichte	1/ 1	-/ -	-/ -	8/ 8	-/ -	-/ -	-/ -	9/ 9
Verdrängung	3/ 3	-/ -	2/ 8	-/ -	-/ -	-/ -	1/ 3	6/ 14
Gesamt	17/17	7/ 7	26/106	12/12	1/ 1	1/ 3	1/ 3	66/149

*Anmerkungen.* NV, naive Vorstellung; ZV, Zwischenvorstellung; WV, wissenschaftliche Vorstellung; <sup>a</sup>Hierunter sind Items subsumiert, die ein naive gegen eine wissenschaftliche Vorstellung testen. <sup>b</sup>Anzahl vorgegebenen Antwortalternativen; <sup>c</sup>Anzahl Aufgabenstämme/Anzahl Items;

#### 4.3.1.2 Weitere Erhebungsinstrumente in Studie 1

Neben dem *Naturwissenschaftlichen Wissen* wurden in Studie 1 einige Kontrollvariablen erhoben: Das Textverständnis mit dem Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6; Untertest Textverständnis; Lenhard & Schneider, 2006), die allgemeine Intelligenz anhand zweier Untertests des Grundintelligenztests Skala 2 – Revision (CFT 20-R; Untertests Topologien und Reihenfortsetzen; Weiß, 2006) sowie naturwissenschaftsbezogene Einstellungen und Selbstkonzept (vgl. Bos et al., 2005). Zusätzlich wurde der sozioökonomische Status über einen Elternfragebogen (vgl. Bos et al., 2005) sowie demographische Schülervariablen und Informationen über vorangegangenen Unterricht in den untersuchten Themenbereichen über einen Lehrerfragebogen erhoben. Die Kontrollvariablen finden in den Auswertungen der vorliegenden Arbeit allerdings keine Berücksichtigung

#### 4.3.1.3 Stichprobe der Studie 1

Für die Teilnahme an der Studie wurden insgesamt 87 Grundschulklassen im Münsterland und im Großraum München akquiriert, die sowohl aus ländlichen wie auch städtischen Regionen

stammten. Einzige Voraussetzung für die Teilnahme war die Jahrgangsstufe der Klassen (2., 3. oder 4. Jahrgangsstufe der Grundschule). Da die Stichprobe aus freiwillig zur Teilnahme bereiten Klassen gebildet wurde, handelt es sich um eine sog. Gelegenheitsstichprobe.

Insgesamt nahmen N=1915 Schüler am Ende des Schuljahres 2009/2010 an der Studie teil, von denen N=1820 (918 Jungen, 894 Mädchen, 8 fehlende Angaben) für die Analysen der vorliegenden Arbeit berücksichtigt wurden<sup>4</sup>. Wie aus Tabelle 3 ersichtlich, waren davon N=598 der zweiten, N=608 der dritten und N=614 der vierten Jahrgangsstufe zuzuordnen. Die Schüler waren zum Zeitpunkt der Studie durchschnittlich 9.20 Jahre alt, wobei das durchschnittliche Alter mit der Klassenstufe anstieg (vgl. Tabelle 3). Während in der vierten Klassenstufe die Anzahl an Jungen und Mädchen ausgeglichen war, nahmen in der zweiten Klasse mehr Mädchen, in der dritten Klasse mehr Jungen an der Studie teil. Die durchschnittliche Anzahl von Schülern in einer Klasse betrug 21 Kinder. Darüber hinaus sind in Tabelle 3 Kennwerte der Stichprobe zu einigen der erhobenen Kontrollvariablen angegeben.

**Tabelle 3 Stichprobenbeschreibung**

	N <sub>Klassen</sub>	N <sub>Schüler</sub>	Alter <sup>b</sup>	Geschlecht <sup>c</sup>	ELFE <sup>d</sup>	CFT <sup>d</sup>	Sprache <sup>e</sup>	Bücher <sup>f</sup>
Klasse 2	28	598	8.14 (0.51)	275/323	8.86	12.37	69.4/26.6	3.72
Klasse 3	29	608	9.26 (0.48)	338/267/3	13.56	15.02	74.8/20.4	4.06
Klasse 4	30	614	10.28 (0.47)	305/304/5	16.19	16.30	71.2/24.1	4.19
Gesamt	87	1820	9.20 (0.99)	918/894/8	12.89	14.57	71.8/23.7	3.99

*Anmerkungen.* <sup>a</sup>In dieser Tabelle werden nur die Werte der Schüler dargestellt, die tatsächlich das Testheft *Naturwissenschaftliches Wissen* bearbeitet haben. <sup>b</sup>M(SD) <sup>c</sup>Anzahl Jungen/Anzahl Mädchen/Anzahl fehlende Angaben <sup>d</sup>M der Anzahl richtig gelöster Aufgaben <sup>e</sup>Schülerangabe zum häuslichen Sprachgebrauch in Prozent: zu Hause wird nur deutsch gesprochen/zu Hause wird noch eine andere Sprache gesprochen (fehlende Prozente sind fehlende Angaben) (vgl. Lenhard & Schneider, 2006) <sup>f</sup>Mittelwert der Schülerangabe zum familiären Buchbesitz: fünfstufige Likert-Skala mit den Endpunkten „keine oder nur sehr wenige Bücher (0-10)“ bis „genug, um drei oder mehr Regale zu füllen (über 200)“ (vgl. Bos et al., 2005)

Durch das zur Zusammenstellung der Testhefte verwendete Rotationsdesign ergibt sich je Aufgabe eine durchschnittliche Stichprobengröße von N=603 Kinder. Dieser Wert schwankte allerdings zwischen der am seltensten bearbeiteten Aufgabe mit N=159 und der am häufigsten bearbeiteten

<sup>4</sup> Die Schüler der Stichprobe, die nicht die Aufgaben zu den untersuchten Themenbereichen bearbeiteten, z.B. weil sie an dem entsprechenden Testtag aus Krankheitsgründen fehlten, wurden bei den Auswertungen nicht berücksichtigt.

Aufgabe mit  $N=1816$ . Aus diesen Werten ergibt sich, dass je Aufgabe im Mittel 66.88% der Werte fehlen (Min=0.22%, Max=91.26%).

#### **4.3.1.4 Durchführung der Studie 1**

Die teilnehmenden Schüler wurden in drei Sitzungen von je einer Unterrichtsdoppelstunde (ca. 90 Minuten) im Klassenverband mit den schriftlichen Tests befragt. In den ersten beiden Sitzungen wurden dabei die Tests zum *Naturwissenschaftlichen Wissen* und zum *Wissen über Naturwissenschaften*<sup>5</sup> in alternierender Reihenfolge durchgeführt. Die Bearbeitung der Fragebögen war dabei von einem geschulten Testleiter angeleitet, der alle Aufgabenstellungen und Antwortalternativen zur Minimierung des Einflusses des Leseverständnisses vorlas und bei 26 der 66 Aufgaben zum *Naturwissenschaftlichen Wissen* die im Aufgabenstamm dargestellte Situation mit realen Gegenständen demonstrierte. Zusätzlich wurde die Durchführung von einer Projektion der Aufgaben sowie einem Testassistenten, der für die Beantwortung von Fragen der Schüler zur Aufgabebearbeitung verantwortlich war, begleitet.

Die Erfassung der Kontrollvariablen fand in der dritten Sitzung im Klassenverband statt. Der Fragebogen zum sozioökonomischen Status wurde zum Ende der Studie über die Schüler an die Eltern ausgegeben.

### **4.3.2 Datenaufbereitung und –auswertung in Studie 1**

#### **4.3.2.1 Kodierung der Schülerantworten auf die Aufgaben in Studie 1**

Die Auswertungen erfolgten auf Grundlage der, teilweise in den Aufgabenstämmen genesteten, 149 Items. Alle Items wurden dichotom kodiert. Bei FC- und MC-Aufgaben wurde dabei jeweils die Auswahl des höchsten vorgegebenen Niveaus bepunktet, d.h. mit 1 kodiert. Im Falle von MS-Aufgaben wurde bei Antwortalternativen auf dem naiven Niveau die Ablehnung, bei Antworten auf dem mittleren und wissenschaftlichen Niveau die Zustimmung bepunktet. Bei den drei Items, bei denen ein graphischer Output erwartet wurde, wurde eine Lösung, die das Verständnis des wissenschaftlichen Niveaus anzeigte, als korrekte Lösung bewertet und mit 1 kodiert. Die Vorstellungen der Schüler in ihren Antworten auf die offene Aufgabe wurden von drei geschulten Kodierern anhand eines Kodierleitfadens, der auf Grundlage der *Construct Map* entwickelt worden war, den drei Kompetenzniveaus zugeordnet. Die Interrater-Reliabilität, die anhand einer Substichprobe von  $N=546$  Schülern überprüft wurde, lag bei Spearman's  $\rho=0.74$  und war damit zufriedenstellend. Aus den Niveauzuordnungen wurden je Schüler drei dichotom kodierte Variablen für die offene Aufgabe erstellt: Für das untere Niveau wurde das Nicht-Vorliegen, für das Zwischen- und wissenschaftliche Niveau das Vorliegen einer entsprechenden Vorstellung bepunktet.

---

<sup>5</sup> Für diese Dimension wurden im Projekt Science-P ebenfalls Instrumente entwickelt. Diese Daten wurden für die vorliegende Arbeit nicht berücksichtigt.



#### 4.3.2.2 Kodierung der Aufgaben hinsichtlich der untersuchten Merkmale in Studie 1

Zur Untersuchung der Fragestellungen wurden alle Items hinsichtlich der beschriebenen Merkmale kodiert. Dabei wurde jedes Item nur einer Ausprägung des jeweiligen Merkmals zugewiesen. War die Zuordnung eines Items zu keiner Merkmalsausprägung eindeutig möglich, wurde ein fehlender Wert kodiert.

Die meisten Merkmale wurden dichotom kodiert (vgl. Tabelle 1). Drei Merkmale (Antwortformat, Temperatur, Kongruenz) wurden nominal mit drei bis fünf Stufen kodiert. Für die Textlänge wurde die Anzahl der Wörter im Aufgabenstamm ausgezählt und als intervallskalierte Kodierung angegeben.

In den meisten Fällen war eine niedrig-inferente Kodierung der Merkmale, also eine Zuweisung von Merkmalsausprägungen ohne Interpretation, möglich. So ist es z.B. eindeutig bestimmbar, ob ein Phänomen drinnen oder draußen situiert ist. Hinsichtlich einiger Merkmale erforderte die dichotome Kategorisierung allerdings eine Interpretationsleistung. Beispielsweise ist nicht ohne weiteres bestimmbar, ob ein Gegenstand groß oder klein ist bzw. ab wann ein Gegenstand von den Probanden nicht mehr als klein, sondern als groß eingestuft bzw. wahrgenommen wird. Dieses Problem betrifft die Merkmale Größe, Gewicht und Hohl, die für die Komponente Auftrieb untersucht werden sollten.

Da das Wirken der Aufgabenmerkmale über assoziative Prozesse bei der Bearbeitung das Antwortverhalten beeinflusst, ist hinsichtlich der Merkmale keine normativ richtige Zuordnung der Items notwendig. Hingegen ist es von Bedeutung, wie die Wahrnehmung der Probanden bezüglich der Merkmale ist. Daher wurde zunächst eine Einteilung der Aufgaben nach der Größe und dem Gewicht vorgenommen, bei der versucht wurde, subjektiv die Perspektive von Grundschulern auf die beurteilten Objekte zu berücksichtigen.

Um diese Kategorisierung zu validieren, wurden anschließend  $N=4$  Schüler der zweiten bis fünften Klassenstufe<sup>6</sup> in einer Expertenvalidierung gebeten, ihrerseits die Gegenstände in Bezug auf die fraglichen Merkmale zu beurteilen. Die mittlere paarweise Übereinstimmung mit der von mir vorgenommenen Einteilung lag für das Merkmal Größe bei  $C=.57^7$ , für Gewicht bei  $C=.61$  und für das Merkmal Hohl bei  $C=.41$ . Eine Überprüfung der paarweisen Übereinstimmung der Schüler untereinander ergab  $C_{\text{Größe}}=.64$ ,  $C_{\text{Gewicht}}=.71$  und  $C_{\text{Hohl}}=.64$ . Eine genaue Inspektion der Urteile der Schüler zeigte, dass es allerdings nur drei Fälle insgesamt gab, in denen alle Schüler nicht mit der

---

<sup>6</sup> Da die teilnehmenden Schüler der Expertenvalidierung sich noch am Schuljahresanfang der jeweiligen Klassenstufe befanden und die Probanden der Querschnittstudie zum Schuljahresende befragt wurde, ist somit eine annähernde Altersgleichheit zur Stichprobe der Querschnittstudie gegeben.

<sup>7</sup> Für die Berechnung der Übereinstimmung der Experten bei der Merkmalsbewertung wurde der Kontingenzkoeffizient  $C$  berechnet. Bei der Beurteilung der Werte ist zu berücksichtigen, dass der Wertebereich von  $C$  begrenzt ist und sich mit steigender Anzahl von Kategorien 1 annähert. Da im vorliegenden Fall nur 2 bzw. 3 Kategorien betrachtet wurden ist  $C_{\text{max}}=.71$  bzw.  $C_{\text{max}}=.82$ .

für die Auswertungen vorgenommenen Einteilung übereinstimmten. In diesen Fällen wurde eine Anpassung der Kodierung der Aufgabenmerkmale vorgenommen. Konkret betroffen waren drei Beurteilungen des Merkmals hohl bei den Objekten Ente, Apfel und Nilpferd. Diese wurden aufgrund einzelner Erfahrungen mit Schülern in der vorgehenden Validierungsstudie (siehe Kapitel 4.3.1.1), die diese explizit als hohl beurteilten, entsprechend klassifiziert. Die Teilnehmer der Expertenvalidierung schätzten diese drei Objekte allerdings übereinstimmend als ausgefüllt ein. Eine Anpassung der Kodierungen führte zu einer Verbesserung der mittleren paarweisen Übereinstimmung bei diesem Merkmal auf  $C_{\text{Hohl}} = .66$ . Die Übereinstimmung zwischen meiner Einstufung und den Einschätzungen der Schüler liegt damit im gleichen Bereich wie die Übereinstimmung der Schüler untereinander.

Die Zuweisung der Items zu den postulierten Kompetenzniveaus wurde anhand des höchsten erreichbaren Niveaus vorgenommen. So wurde z.B. eine MC-Aufgabe, in der alle drei Niveaus als Alternativen angeboten wurden, dem höchsten Niveau der wissenschaftlichen Vorstellungen zugeordnet. Ein MS-Item, in dem die Zustimmung oder Ablehnung zu einer Naiven Vorstellung erfragt wurde, wurde dem untersten Niveau zugeordnet. Welchem der Niveaus die Antwortalternativen zugeordnet waren, war zuvor durch die Operationalisierung der Konstrukte in der *Construct Map* festgehalten worden (siehe Kapitel 4.3.1.1). Eine Expertenvalidierung dieser Zuordnung ergab eine mittlere Übereinstimmung von  $N=12$  Experten von  $r=.77$  mit der im Projekt vorgenommenen Zuweisung. Diese lag im Themenbereich Schwimmen/Sinken mit  $r=.72$  etwas niedriger als im Themenbereich Verdunstung/Kondensation mit  $r=.86$  und erreichte somit zufriedenstellende Werte. Es gab keinen Fall, in dem keiner der Experten mit der Projekteinteilung übereinstimmte. Daher wurde keine Veränderung der Kodierung der Aufgaben hinsichtlich der Niveaus vorgenommen.

#### 4.3.2.3 *Dimensionalitätsprüfung, Itemselektion und zentrale psychometrische Kennwerte der Skalen und Items in Studie I*

Da mit einer Erweiterung des linear-logistischen Testmodells zur Überprüfung der Hypothesen eine Weiterentwicklung des Rasch-Modells (Rasch, 1960) gerechnet werden sollte (siehe 4.3.2.4), stand vor dem Beginn der eigentlichen Analysen eine Prüfung und Selektion der Items hinsichtlich ihrer Passung an dieses Modell, die im Folgenden beschrieben wird. Die Kontrolle der psychometrischen Eigenschaften der Items, die Selektion geeigneter Items und die Prüfung der Dimensionalität der eingesetzten Aufgaben wurden mit dem Programm ACER Conquest 2.0 (Wu, Adam & Wilson, 2005) durchgeführt.

Zur Prüfung der Dimensionalität wurden ein- und mehrdimensionale einparametrische logistische Modelle (1-PL Modelle) an die Daten angepasst und eine Entscheidung für eine dimensionale Struktur, die den weiteren Auswertungen zugrunde liegt, anhand des Chi-Quadrat-Test getroffen (Andersen, 1973). Ein Vergleich eines ein- und eines zweidimensionalen Modells mit den

Dimensionen Schwimmen/Sinken und Verdunstung/Kondensation mittels des Chi-Quadrat-Tests offenbarte eine signifikant bessere Passung des zweidimensionalen Modells ( $\chi^2(2)=246.83$ ,  $p<0.001$ ).

Die psychometrischen Eigenschaften der Items wurden daher anhand der jeweils eindimensionalen Skalierungen der beiden Themenbereiche überprüft. Die Auswahl von Items erfolgte anhand der gewichteten Abweichungsquadrate (*weighted mean square*, MNSQ; Ausschluss wenn  $MNSQ > 1.30$ ), der Trennschärfe (Ausschluss wenn  $r_{it} < .10$ ) und dem Vorliegen von Boden- und Deckeneffekten (Ausschluss wenn Anzahl korrekter Antworten  $< .10 | > .90$ ).

Für den Themenbereich Verdunstung/Kondensation wurden insgesamt acht Items, drei aufgrund mangelnder Trennschärfe und fünf wegen Bodeneffekten, aus den folgenden Analysen ausgeschlossen. Von den ausgeschlossenen Items waren zwei der Komponente Verdunstung und sechs der Komponente Kondensation zuzuordnen. Bei erneuter Skalierung der verbleibenden 66 Items ergab sich eine Reliabilität der Personenparameterschätzung (*Warms weighted likelihood estimates*, WLE) von 0.60 bei einer Varianz von 0.48.

Insgesamt dreizehn Items wurden bei Schwimmen und Sinken ausgeschlossen. Von diesen hatten neun eine ungenügende Trennschärfe, eines einen ungenügenden Fit, ein Item einen Boden- und zwei Items einen Deckeneffekt. Mit zwölf der dreizehn Items war die Mehrheit der auszuschließenden Items der Komponente Auftrieb zuzuordnen. Ein Item der Komponente Verdrängung wurde eliminiert. Nach Ausschluss dieser Items ergab sich für die restlichen 62 Items eine Reliabilität von 0.77 bei einer Varianz von 1.00.

Eine Übersicht der eingesetzten Aufgaben und Items der Querschnittstudie sowie eine vollständige Darstellung von Kennwerten für die verbleibenden Items sind in Anhang A02 bzw. Anhang A03 zu finden.

#### 4.3.2.4 Vorgehen und Verfahren zur Hypothesenprüfung in Studie 1: Generalisierte lineare gemischte Modelle mit gekreuzten zufälligen Effekten

Da an die Daten dieser Studie bereits 1-PL Modelle angepasst worden sind, könnten die Schwierigkeitsparameter der Items direkt als unabhängige Variablen in varianz- oder regressionsanalytische Verfahren zur Überprüfung des Einflusses von Aufgabenmerkmalen eingeführt werden (vgl. 2.3.1). In der vorliegenden Arbeit wurde allerdings die Formulierung probabilistischer Modelle angestrebt, in denen die Skalierung der Daten simultan mit der Überprüfung des Effektes der Aufgabenmerkmale vorgenommen wird.

Ausgangspunkt für die Formulierung solcher Modelle ist häufig das 1-PL Modell (Rasch, 1960):

$$P(X_{ij} = 1 | \theta_j, \sigma_i) = \frac{\exp(\theta_j - \sigma_i)}{1 + \exp(\theta_j - \sigma_i)}$$

In diesem Modell wird für jede Person ein Parameter  $\theta_j$  und für jedes Item ein Parameter  $\sigma_i$  geschätzt. Somit wird die Wahrscheinlichkeit einer richtigen Antwort ( $X_{ij} = 1$ ) unter der

Bedingung einer bestimmten Fähigkeits- und einer bestimmten Schwierigkeitsausprägung geschätzt.

Eine Weiterentwicklung dieses Modells stellt das linear-logistische Testmodell (LLTM; Fischer, 1973) dar. In diesem Modell wird die Itemschwierigkeit als lineare Kombination von Prädiktoren modelliert:

$$P(X_{ij} = 1 | \theta_j, q_{ik}, \eta_k) = \frac{\exp(\theta_j - (\sum_{k=1}^K q_{ik} \eta_k))}{1 - \exp(\theta_j - (\sum_{k=1}^K q_{ik} \eta_k))}$$

Dabei werden Parameter der Items durch die Itemkovariaten  $q_{ik}$  und die dazu gehörigen Gewichte  $\eta_k$  dargestellt.

Die Behandlung von Items als zufällig, im Gegensatz zu festen Effekten, hat mehrere Gründe (vgl. De Boeck, 2008): Die Items können als eine zufällige Ziehung aus einem gedachten Universum möglicher Items zur Erfassung der untersuchten Kompetenzen betrachtet werden. Diese Überlegung wird unterstützt durch die Verteilung von Items auf verschiedene Testhefte mittels Rotationsdesign sowie die Aussonderung von Items während der Instrumententwicklung. Das LLTM macht die Annahme, dass die Varianz der Itemschwierigkeiten perfekt durch die untersuchten Prädiktoren aufzuklären ist. Diese Annahme ist sowohl theoretisch wie auch praktisch unrealistisch. Daher wird das LLTM für die vorliegende Arbeit um einen zufälligen Fehlerterm für Items erweitert (LLTM+e; De Boeck, 2008; Janssen, Schepers & Peres, 2004). Das LLTM+e hat allgemein folgende Form:

$$P(X_{ij} = 1 | \theta_j, q_{ik}, \eta_k) = \frac{\exp(\theta_j - (\sum_{k=1}^K q_{ik} \eta_k + \varepsilon_i))}{1 - \exp(\theta_j - (\sum_{k=1}^K q_{ik} \eta_k + \varepsilon_i))}$$

Die lineare Kombination von festen Prädiktoren zur Modellierung der Itemschwierigkeit wurde um den zufälligen Term  $\varepsilon_i$  erweitert.

In der vorliegenden Arbeit wurden die Modelle als generalisierte lineare gemischte Modelle mit dem Paket lme4 in R (*generalized linear mixed models*, GLMMs; Doran, Bates, Bliese & Dowling, 2007; Baayen, Davidson & Bates, 2008; De Boeck et al., 2011; De Boeck & Wilson, 2004) berechnet. In Modellen dieser Art ist es möglich, sowohl für die Personen- wie auch die Aufgabenstichprobe einen zufälligen Effekt einzuführen (gekreuzte zufällige Effekte; *crossed random effects*). Um der Mehrebenenstruktur der Stichprobe auf der Personenseite gerecht zu werden, wurde neben einem zufälligen Effekt für einzelne Personen und Items ein zufälliger Effekt für die Schulklassen, in denen die Schüler für die vorliegende Studie untersucht wurden, in das Modell aufgenommen. Die allgemeine Formulierung der untersuchten Modelle hatte in meiner Untersuchung demnach folgende Form:

$$P(X = 1) = q_{i1}\eta_1 + q_{i2}\eta_2 + \dots + q_{ik}\eta_k + \varepsilon_j + \varepsilon_i + \varepsilon_{class}$$

Die Regressionsgewichte sind dabei als sog. *Leichtigkeitsparameter* zu interpretieren: Die Gewichte geben an, um welchen Betrag sich die Wahrscheinlichkeit der Lösung eines Items bei Vorliegen des jeweiligen Merkmals erhöht. Die Signifikanz der Koeffizienten für die untersuchten

Aufgabenmerkmale kann anhand der z-Werte getestet werden (De Boeck et al., 2011). Für mögliche Modellvergleiche sollten unterschiedliche Kriterien, wie das *bayesian information criterion* (BIC; Schwartz, 1978) und der Chi-Quadrat-Test (Andersen, 1973; nur geeignet für genestete Modelle), herangezogen werden.

Die Modelle wurden mit einem allgemeinen Y-Achsenabschnitt  $\eta_1$  (overall intercept) modelliert. Dazu wurde bei der Berechnung der Modelle eine Kategorie der untersuchten Prädiktoren als Referenzkategorie genutzt und der Mittelwert dieser Kategorie geschätzt. Das so ermittelte Regressionsgewicht wurde auf eine signifikante Unterschiedlichkeit zu Null getestet. Fällt dieser Test allerdings signifikant aus, ist davon nicht auf eine signifikante Unterschiedlichkeit der Schwierigkeit von Items mit verschiedener Merkmalsausprägung zu schließen. Man kann hier lediglich die Aussage treffen, dass die entsprechenden Items im Mittel leichter (bei positiven Koeffizienten) oder schwerer (bei negativem Koeffizienten) waren als die mittlere Personenfähigkeit. Daher wird bei der Ergebnisdarstellung (siehe Kapitel 4.4.3) nicht auf signifikante Y-Achsenabschnitte eingegangen. Die Regressionsgewichte  $\eta_k$  hingegen weisen bei Signifikanz auf unterschiedliche Itemschwierigkeiten hin. Sie entsprechen der Mittelwertsdifferenz zwischen Referenz- und Testkategorie. Weichen diese Differenzen signifikant von Null ab, sind die Schwierigkeitsunterschiede zwischen Items substantiell: Items der Testkategorie sind leichter (bei positivem Koeffizienten) oder schwerer (bei negativem Koeffizienten) als Items der Referenzkategorie.

Die beschriebenen Modelle bieten den Vorteil großer Flexibilität für die Art der möglichen zu untersuchenden Kovariaten: Es ist möglich gleichzeitig Merkmale von Personen und Aufgaben als Prädiktoren zu berücksichtigen und es kommen dichotome, kategoriale und kontinuierliche Variablen als Prädiktoren in Frage.

Es wäre außerdem möglich, den zufälligen Effekt für Items in weitere Parameter zu zerlegen. So wäre es z.B. möglich Varianzkomponenten für unter einem Aufgabenstamm genestete Items einzuführen, um so einer möglichen Abhängigkeit der Items gerecht zu werden, oder für unterschiedliche Gruppen von Items eine heteroskedastische Fehlervarianz zu modellieren (De Boeck et al., 2011). Auf diese beiden Möglichkeiten wird in der vorliegenden Arbeit allerdings verzichtet, da die Regressionsgewichte für Itemprädiktoren sich in solchen Modellen auf unterschiedliche Varianzkomponenten bezögen und somit nicht mehr im Sinne der Fragestellungen interpretierbar wären.

Neben der Flexibilität der Modelle bieten diese außerdem einen vorteilhaften Umgang mit fehlenden Daten. Da die Daten der vorliegenden Studie mit einem Rotationsdesign der Testhefte erhoben wurden, kam, neben dem vereinzelt Fehlen von Werten im eigentlichen Datensatz, ein großer Anteil fehlender Daten hinzu, die design-bedingt entstanden sind. Diese Voraussetzung führte dazu, dass Imputationsmethoden zur Ersetzung der fehlenden Werte in diesem Fall nicht geeignet waren. In den generalisierten linearen gemischten Modellen, die für die Auswertung der

Daten gerechnet wurden, wurden die Items unter die Personen geschachtelt. D.h. dass jeder einzelne Datenpunkt, der sich aus einer Kombination einer Person und eines Items ergab, einen Fall in der den Analysen zugrunde liegenden Datenmatrix darstellte. Kam es zum Ausschluss von Datenpunkten, weil z.B. eine Person ein Item nicht bearbeitet hat, wurde kein listenweiser oder paarweiser Fallausschluss angewandt, der zur Verzerrung der Statistiken führen könnte (Lüdtke, Robitzsch, Trautwein & Köller, 2007), sondern nur genau dieser Datenpunkt stand für die Analysen nicht zu Verfügung. Eine systematische Abweichung der Schätzungen durch unangemessenen Umgang mit fehlenden Werten war also nicht zu befürchten.

Zur Untersuchung der Fragestellungen mit inferenzstatistischen Methoden habe ich in der vorliegenden Arbeit folgendes Vorgehen angewandt: Zunächst wurde der Einfluss jedes Aufgabenmerkmals auf die Aufgabenschwierigkeiten separat mit dem LLTM+e untersucht. Anschließend wurde je Kontextmerkmal ein Modell mit nur den Haupteffekten des jeweiligen Merkmals und der Niveaus und ein Modell, das neben den Haupteffekten die Interaktion zwischen Merkmal und Niveaus beinhaltet, berechnet. Diese beiden Modelle wurden dann mit dem Chi-Quadrat-Test verglichen, um so mögliche signifikante Interaktionen zu finden. Tabelle 4 und Tabelle 5 geben einen Überblick darüber, welche Modelle je (Themen-) Bereich berechnet wurden.

**Tabelle 4 Übersicht über inferenzstatistische Auswertungen in Studie 1 für den Themenbereich Verdunstung/Kondensation**

	Schritt 1			Schritt 2		
	Untersuchung von Haupteffekten			Untersuchung von Interaktionen		
	VuK	Vd	K	VuK	Vd	K
formale Aufgabenmerkmale						
Antwortformat	✓	✓	✓			
Demonstration mit realen Gegenständen	✓	✓	✓			
Textlänge	✓	✓	✓			
Merkmale aus Kompetenzmodell						
Niveaus	✓	✓	✓			
Basiskonzept	✓	✗	✗			
Kontextmerkmale						
Ort	✓	✓	✓	✓	✓	✗
Temperatur	✓	✓	✓	✓	✓	✓
Wasserspiegel	✗	✓	✗	✗	✓	✗
Kondensationskern	✗	✗	✓	✗	✗	✗
Quelle	✗	✗	✓	✗	✗	✓

*Anmerkungen.* VuK, Verdunstung und Kondensation; Vd, Verdunstung; K, Kondensation; ✓, Modell wird berechnet; ✗, Modell wird nicht berechnet; Die Untersuchung von Interaktionen wird, entsprechend der Fragestellungen, nur für die Kontextmerkmale angestrebt.

In den folgenden Abschnitten werden die Ergebnisse der durchgeführten Analysen vorgestellt. Dabei werden im Rahmen einer Aufgabenanalyse zunächst Befunde zu der Verteilung der Aufgaben hinsichtlich der untersuchten Merkmale vorgestellt (siehe Kapitel 4.4.1), bevor anschließend auf die deskriptiven Ergebnisse eingegangen wird (siehe Kapitel 4.4.2). Anschließend

werden die Ergebnisse der inferenzstatistischen Analysen mit Bezug auf die Fragestellungen berichtet (siehe Kapitel 4.4.3). Die Beschreibung der Ergebnisse zu Verdunstung/Kondensation ist jeweils der zu Schwimmen/Sinken voran gestellt. Innerhalb der Abschnitte zu den Themenbereichen werden zunächst Ergebnisse über alle Komponenten hinweg berichtet, bevor auf die einzelnen Aspekte eingegangen wird.

**Tabelle 5 Übersicht über inferenzstatistische Auswertungen in Studie 1 für den Themenbereich Schwimmen/Sinken**

	Schritt 1				Schritt 2			
	Untersuchung von Haupteffekten				Untersuchung von Interaktionen			
	SuS	A	D	V	SuS	A	D	V
	formale Aufgabenmerkmale							
Antwortformat	✓	✓	✓	✓				
Demonstration mit realen Gegenständen	✓	✓	✗	✓				
Textlänge	✓	✓	✗	✓				
	Merkmale aus Kompetenzmodell							
Niveaus	✓	✓	✗	✓				
Basiskonzept	✓	✗	✗	✗				
	Kontextmerkmale							
Größe	✗	✓	✗	✗	✗	✓	✗	✗
Gewicht	✗	✓	✗	✗	✗	✓	✗	✗
Hohl	✗	✓	✗	✗	✗	✓	✗	✗
Kongruenz	✗	✗	✓	✗	✗	✗	✗	✗
Vergleich	✗	✗	✗	✓	✗	✗	✗	✗

*Anmerkungen.* SuS, Schwimmen und Sinken; A, Auftrieb; D, Dichte; V, Verdrängung; ✓, Modell wird berechnet; ✗, Modell wird nicht berechnet; Die Untersuchung von Interaktionen wird, entsprechend der Fragestellungen, nur für die Kontextmerkmale angestrebt.

## 4.4 Ergebnisse von Studie 1

### 4.4.1 Ergebnisse der Aufgabenanalyse von Studie 1

#### 4.4.1.1 Verdunstung und Kondensation

In den Abschnitten zur Aufgabenanalyse werden die Items der Studie bzw. ihre Anzahl als abhängige Variablen behandelt. In Tabelle 6 ist die Verteilung der Items zu Verdunstung und Kondensation hinsichtlich der unterschiedlichen Ausprägungen der Aufgabenmerkmale dargestellt.

**Tabelle 6 Verteilung der Items zu Verdunstung und Kondensation auf die Aufgabenmerkmale, Itemmittelwerte und Standardabweichungen**

	Verdunstung		Kondensation		Gesamt	
	N	M/SD	N	M/SD	N	M/SD
Alle Items	36	.61/.18	30	.55/.19	66	.58/.18
	formale Aufgabenmerkmale					
Antwortformat						

<i>forced choice</i>	6	.70/.15	9	.45/.09	15	.55/.17
<i>multiple select</i>	26	.58/.19	20	.60/.20	46	.59/.19
<i>multiple choice</i> <sup>a</sup>	4	.68/.15	1	.45/ –	5	.63/.16
Demonstration mit realen Gegenständen						
nein	32	.60/.18	22	.52/.18	54	.57/.18
ja	4	.72/.13	8	.62/.18	12	.65/.17
Merkmale aus Kompetenzmodell						
Niveaus						
NV	13	.60/.22	9	.59/.24	22	.59/.22
ZV	10	.65/.19	7	.53/.16	17	.60/.18
WV	13	.60/.13	14	.53/.17	17	.56/.15
Kontextmerkmale						
sichtbarer Wasserspiegel <sup>b</sup>						
ja	11	.57/.19	-	-	-	-
nein	25	.63/.17				
Ort						
drinnen	24	.65/.18	25	.56/.19	49	.61/.18
draußen	12	.54/.17	5	.48/.18	17	.52/.17
Temperatur <sup>c,d</sup>						
warm	7	.55/.22	13	.52/.19	28	.53/.18
Sonne	8	.54/.13	-	-	-	-
kalt	20	.65/.17	16	.57/.18	36	.61/.18
Kondensationskern <sup>d,f</sup>						
Fläche	-	-	26	.56/.19	-	-
Dampf			3	.40/.05		
Quelle <sup>d,f</sup>						
vorhanden	-	-	21	.58/.19	-	-
nicht vorhanden			8	.46/.14		

*Anmerkungen.* NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>a</sup>Für die Berechnung der deskriptiven Statistiken und alle weiteren Analysen wurden *forced choice*-Aufgaben, die eine naive einer Zwischen- bzw. eine Zwischen- einer wissenschaftlichen Vorstellungen gegenüberstellen, sowie *multiple choice*-Aufgaben mit 3 oder 4 Antwortalternativen zusammen gefasst. <sup>b</sup>Die Unterscheidung zwischen einem sichtbaren und unsichtbaren Wasserspiegel wurde nur bei Verdunstung vorgenommen. <sup>c</sup>Für Kondensation wurde nur zwischen kalten und warmen Situationen unterschieden, während bei Verdunstung noch Situationen, in denen die Sonne explizit erwähnt wurde, von warmen Situationen abgegrenzt wurden. <sup>d</sup>Da einzelne Items keiner Kategorie des Merkmals eindeutig zuzuordnen waren, werden diese bei diesen Auswertungen nicht berücksichtigt. <sup>f</sup>Die Kontextmerkmale Kondensationskern und Quelle wurden nur für den Bereich Kondensation untersucht.

Das Antwortformat *multiple select* wurde in diesem Themenbereich am häufigsten eingesetzt (n=46, vgl. Tabelle 6). Die große Häufigkeit solcher Items ergab sich, da bei diesem Format mit einem relativ geringen Aufwand an Testzeit mehrere Antwortreaktionen von den Probanden verlangt werden konnten, während bei den anderen Formaten je Item mehr Testaufwand erforderlich war. Eine Demonstration der im Aufgabenstamm dargestellten Situation mit realen Gegenständen war für dieses Thema nur bei wenigen Aufgaben (n=12) und hauptsächlich im Bereich Kondensation (n=8) möglich. Als weiteres formales Aufgabenmerkmal wurde die Textlänge, operationalisiert als Anzahl der im Aufgabentext verwendeten Wörter, betrachtet. Diese Größe schwankte für den Themenbereich zwischen Min=16 und Max=74 Wörtern (M=42.85, SD=13.89), wobei sie für Verdunstung zwischen Min=22 und Max=71 Wörtern (M=41.56, SD=13.49) und für Kondensation zwischen Min=16 und Max=74 Wörtern (M=44.40, SD=14.43)



lag. Die Aufgaben zu Kondensation waren also durch einen etwas längeren Text begleitet als die zu Verdunstung.

Bei der Betrachtung der durch das Kompetenzmodell vorgegebenen Unterscheidung dreier Niveaus fällt auf, dass für beide Bereiche das mittlere Niveau (Verdunstung:  $n=10$ , Kondensation:  $n=7$ ) und für Kondensation auch das untere Niveau ( $n=9$ ) seltener in den Aufgaben repräsentiert waren als das wissenschaftliche Niveau (Verdunstung:  $n=13$ , Kondensation:  $n=14$ ). Dies ist darauf zurück zu führen, dass FC- oder MC-Aufgaben, bei denen die richtige Antwort dem höchsten Niveau entsprach, dem Niveau der *wissenschaftlichen Vorstellungen* zugeordnet wurden, auch wenn als Distraktoren noch Antwortalternativen der anderen Niveaus angeboten wurden.

Von allen Aufgaben zu diesem Themenbereich war der Großteil drinnen situiert ( $n=49$ ), was sich auch an der Verteilung der Items für die einzelnen Komponenten hinsichtlich dieses Kontextmerkmals widerspiegelt. Für Verdunstung waren mehr als doppelt so viele Items drinnen situiert ( $n=24$ ) wie draußen ( $n=12$ ). Für Kondensation war die Verteilung noch extremer: Nur 5 der 30 Items waren draußen situiert.

In Bezug auf die Temperatur war die Unterteilung der Items in warm und kalt für das gesamte Thema etwas ausgeglichener (warm:  $n=28$ , kalt:  $n=36$ ). Bei Betrachtung der beiden Komponenten zeigte sich dies ebenfalls: Für Kondensation waren 13 Items, für Verdunstung 15 Items als warm und 16 bzw. 20 Items als kalt einzustufen.

Für Aufgaben zum Bereich Verdunstung war in mehr als doppelt so vielen Aufgaben der Wasserspiegel unsichtbar ( $n=25$ ) als sichtbar ( $n=11$ ). Bezüglich der Items zur Komponente Kondensation zeigte sich, dass in den meisten Aufgaben eine potentielle Quelle des kondensierten Wassers vorhanden war ( $n=21$ ).

Eine Inspektion der Korrelationskoeffizienten zeigte, dass keines der untersuchten Kontextmerkmale einen signifikanten Zusammenhang zum postulierten Kompetenzniveau aufweist (vgl. Tabelle 7). Wohl aber waren nicht alle Zellen der Kreuztabellen aus jeweils einem Kontextmerkmal und dem Niveau der Items von mindestens einem Item besetzt: Bei Items der Komponente Kondensation wurde keine Zwischenvorstellung als Antwort angeboten, wenn das zu erklärende Phänomen draußen situiert ist und keine naive oder Zwischenvorstellung wenn Wasser als sichtbarer Nebel in der Luft kondensiert.

**Tabelle 7 Zusammenhänge zwischen den untersuchten Merkmalen des Kontexts und dem Kompetenzniveau für den Themenbereich Verdunstung und Kondensation**

		Level	Temperatur	Wasser- spiegel	Konden- sationskern	Quelle
Gesamt	Ort	.28	.69***	-	-	-
	Temperatur	.13	-	-	-	-
Vd	Ort	.37	.77***	.47***	-	-
	Temperatur	.25	-	.66***	-	-

	Wasserspiegel	.25	-	-	-	-
K	Ort	.32	.36	-	.52**	.65***
	Temperatur	.20	-	-	.08	.40*
	Kondensations- kern	.38	-	-	-	.30
	Quelle	.27	-	-	-	-

*Anmerkungen.* Vd, Verdunstung; K, Kondensation; <sup>†</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001; Zur Berechnung der Zusammenhänge wurde aufgrund des nominalen Skalenniveaus der Variablen das Assoziationsmaß Cramér's V, bzw. der Phi-Koeffizient bei zwei dichotomen Variablen, angegeben.

Außerdem korrelierten einige der Merkmale des Kontexts bedeutsam miteinander (vgl. Tabelle 7 sowie Anhang A03.1): Sowohl über den gesamten Themenbereich ( $r=.69$ ,  $p<.001$ ) wie auch nur für die Komponente Verdunstung ( $r=.77$ ,  $p<.001$ ) gab es keine Aufgaben, die drinnen situiert waren und in denen die Sonne als Energiequelle fungierte. Darüber hinaus wurden für Verdunstung keine Aufgaben konstruiert, in denen draußen ein sichtbarer Wasserspiegel vorlag ( $r=.47$ ,  $p<.001$ ). Desweiteren wurde in Aufgaben mit sichtbarem Wasserspiegel ebenfalls nicht die Sonne erwähnt und in solchen Aufgaben war die Temperatur häufiger erhöht, während bei Aufgaben mit unsichtbarem Wasserspiegel häufiger eine Energiequelle für eine höhere Temperatur sorgte ( $r=.66$ ,  $p<.001$ ).

Bei den Items zu Kondensation wurde bei den drinnen situierten Aufgaben häufiger ein Phänomen beschrieben, bei dem Wasser auf einer Fläche und nicht als Dampf kondensierte ( $r=.52$ ,  $p<.001$ ) und bei dem eine potentielle Quelle des Wassers sichtbar war ( $r=.65$ ,  $p<.001$ ). Eine Wasserquelle war außerdem häufiger bei Aufgaben mit erhöhter Temperatur vorhanden, während in Items ohne explizite Energiezufuhr in annähernd gleichvielen Aufgaben eine Quelle vorhanden war oder nicht ( $r=.40$ ,  $p<.05$ ). Die vollständigen Kontingenztabellen zu diesen Befunden sind in Anhang A03.1 zu finden.

#### 4.4.1.2 Schwimmen und Sinken

Wie Tabelle 8 zu entnehmen ist, war auch im Themenbereich Schwimmen/Sinken das Antwortformat *multiple select* insgesamt am häufigsten vertreten ( $n=40$ ). Bezogen auf die Komponenten galt dies allerdings nur für Auftrieb ( $n=33$ ) und Verdrängung ( $n=7$ ). Bei Dichte waren MC-Aufgaben ( $n=8$ ) am häufigsten.

Eine Demonstration der betrachteten Phänomene war bei diesem Thema häufig möglich ( $n=42$ ), allerdings wiederum nur in Bezug auf die Items zu Auftrieb ( $n=30$ ) und Verdrängung ( $n=12$ ); während bei Dichte keine Demonstrationen möglich waren. Die Anzahl der im Aufgabentext verwendeten Wörter bewegte sich für Schwimmen und Sinken zwischen Min=20 und Max=106 ( $M=47.84$ ,  $SD=22.37$ ). Dabei lag diese Zahl für die Komponente Auftrieb zwischen Min=20 und Max=73 ( $M=43.65$ ,  $SD=14.83$ ), für Dichte zwischen Min=23 und Max=35 ( $M=24.33$ ,  $SD=4.00$ ) und für Verdrängung zwischen Min=43 und Max=106 ( $M=77.00$ ,  $SD=20.62$ ). Somit hatten Items

der Komponente Verdrängung die längsten, der Komponente Dichte die kürzesten einführenden Texte.

**Tabelle 8 Verteilung der Items zu Schwimmen und Sinken auf die Aufgabenmerkmale, Itemmittelwerte und Standardabweichungen**

	Auftrieb		Dichte		Verdrängung		Gesamt	
	N	M/SD	N	M/SD	N	M/SD	N	M/SD
Alle Items	40	.62/.16	9	.62/.13	13	.57/.19	62	.61/.26
formale Aufgabenmerkmale								
Antwortformat								
<i>forced choice</i>	5	.55/.11	1	.60/ –	3	.72/.13	9	.60/.13
<i>multiple select</i>	33	.65/.16	-	-	7	.61/.15	40	.64/.15
<i>multiple choice</i> <sup>a</sup>	-	-	8	.62/.14	-	-	8	.62/.14
andere	-	-	-	-	3	.33/.05	3	.33/.05
offen	2	.39/.05	-	-	-	-	2	.39/.05
Demonstration mit realen Gegenständen								
nein	10	.56/.15	9	-	1	.58/ –	20	.59/.14
ja	30	.65/.16	-	-	12	.57/.19	42	.63/.17
Merkmale aus Kompetenzmodell								
Niveaus								
NV	16	.64/.21	-	-	2	.63/.16	18	.64/.20
ZV	10	.61/.14	-	-	-	-	10	.61/.14
WV	14	.62/.10	9	-	11	.56/.20	34	.60/.14
Kontextmerkmale								
Größe <sup>b</sup>								
klein	29	.64/.15	-	-	-	-	-	-
groß	6	.59/.16						
Gewicht <sup>b</sup>								
leicht	30	.64/.15	-	-	-	-	-	-
schwer	10	.57/.19						
Hohl <sup>b</sup>								
gefüllt	26	.64/.14	-	-	-	-	-	-
hohl	14	.60/.19						
Kongruenz <sup>b</sup>								
kongruent	-	-	1	.85/ –	-	-	-	-
gleiches Gewicht			5	.65/.04				
inkongruent			3	.48/.07				
Vergleich <sup>b</sup>								
Einzelurteil	-	-	-	-	4	.43/.20	-	-
Vergleich					9	.64/.15		

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>a</sup>Für die Berechnung der deskriptiven Statistiken und alle weiteren Analysen wurden *forced choice*-Aufgaben, die eine naive einer Zwischen- bzw. eine Zwischen-einer wissenschaftlichen Vorstellungen gegenüberstellen, sowie *multiple choice*-Aufgaben mit 3 oder 4 Antwortalternativen zusammen gefasst. <sup>b</sup>Die Merkmale Größe, Gewicht und Hohl wurden nur für die Komponente Auftrieb untersucht; die Kongruenz nur für Dichte und der Vergleich nur für Verdrängung. <sup>c</sup>Da einzelne Items keiner Kategorie des Merkmals eindeutig zuzuordnen waren, werden diese bei diesen Auswertungen nicht berücksichtigt.

Die Zwischenvorstellungen waren bei diesem Thema, wie auch bei Verdunstung und Kondensation, mit der geringsten Zahl vertreten (n=10). Für die Komponente Auftrieb, bedingt

durch die Häufigkeit von MS-Aufgaben, war allerdings das naive Niveau mit 16 Items sogar häufiger vertreten als das wissenschaftliche mit 14 Items.

Bezüglich der Kontextmerkmale Größe und Gewicht waren die Items zu Auftrieb ungleich verteilt: kleine (n=29) bzw. leichte Gegenstände (n=30) traten häufiger auf als große (n=6) bzw. schwere (n=10). Ebenfalls wurden in mehr Aufgabenstämmen ausgefüllte (n=26) als hohle (n=14) Gegenstände betrachtet.

Bei den Items zu Dichte war nur ein Gegenstandspaar als kongruent einzustufen. In den meisten Aufgaben wurde hingegen eine Größe (Gewicht oder Material) konstant gehalten (n=5) und drei Aufgaben waren als inkongruent einzustufen.

Bei der Verdrängung wurden die Probanden häufiger gebeten Gegenstände miteinander zu vergleichen (n=9) als ein Urteil über einen einzelnen Gegenstand abzugeben (n=4).

**Tabelle 9 Zusammenhänge zwischen den untersuchten Merkmalen des Kontexts und dem Kompetenzniveau für den Themenbereich Schwimmen und Sinken**

		Level	Gewicht	Hohl
A	Größe	.20	.90***	.25
	Gewicht	.08	-	.55**
	Hohl	.06	-	-
D	Kongruenz	-	-	-
V	Vergleich	.28	-	-

*Anmerkungen.* A, Auftrieb; D, Dichte; V, Verdrängung; \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ; Zur Berechnung der Zusammenhänge wurde aufgrund des nominalen Skalenniveaus der Variablen das Assoziationsmaß Cramér's V, bzw. der Phi-Koeffizient bei zwei dichotomen Variablen, angegeben.

Auch für diesen Themenbereich wies kein Kontextmerkmal einen signifikanten Zusammenhang zum postulierten Niveau der Aufgaben auf (siehe Tabelle 9).<sup>8</sup> Für Verdrängung lag allerdings kein Item vor, das einen Vergleich auf naivem Niveau erfragte.

Die Kontextmerkmale der Auftrieb-Items wiesen zum Teil signifikante Korrelationen untereinander auf: Große Gegenstände waren i.d.R. auch schwer und nur selten leicht ( $r = .90$ ,  $p < .001$ ), leichte Gegenstände waren häufiger ausgefüllt als hohl ( $r = .55$ ,  $p < .001$ ). Die Kontingenztabellen zu diesen Ergebnissen sind in Anhang A03.2 zu finden.

<sup>8</sup> Die entsprechende Korrelation konnte für Dichte aufgrund der Invarianz der untersuchten Niveaus bei dieser Komponente nicht berechnet werden.

#### 4.4.2 Deskriptive Ergebnisse von Studie 1

##### 4.4.2.1 Verdunstung und Kondensation

Neben den Häufigkeiten sind in Tabelle 6 die Mittelwerte der Items, die aufgrund der dichotomen Kodierung als relative Lösungshäufigkeiten interpretiert werden können, mit den dazugehörigen Standardabweichungen dargestellt.

Bei den formalen Aufgabenmerkmalen zeigte sich, dass Aufgaben im Format *multiple choice* über beide Komponenten hinweg mit der höchsten Wahrscheinlichkeit gelöst wurden ( $M=.63$ ,  $SD=.16$ ). Bei Verdunstungsitems allein wurden FC-Aufgaben am häufigsten gelöst ( $M=.70$ ,  $SD=.15$ ) und Items in den beiden anderen Formaten seltener gelöst (MC:  $M=.68$ ,  $SD=.15$ ; MS:  $M=.58$ ,  $SD=.19$ ). Bei Kondensation war allerdings die Lösungshäufigkeit für MS-Aufgaben ( $M=.60$ ,  $SD=.20$ ) am höchsten (MC:  $M=.45$ ; FC:  $M=.45$ ,  $SD=.09$ ). Im Mittel ebenfalls häufiger gelöst wurden Aufgaben, in denen das dargestellte Phänomen mit realen Gegenständen demonstriert wurde ( $M=.65$ ,  $SD=.17$ ), im Vergleich zu Aufgaben, bei denen keine Demonstration stattfand ( $M=.57$ ,  $SD=.18$ ). Dieser Unterschied blieb bei separater Betrachtung der beiden Konzepte erhalten (siehe Tabelle 6). Die Betrachtung der Niveaus zeigte, dass Items, die dem Niveau der *Zwischenvorstellungen* zugeordnet waren, am häufigsten gelöst wurden ( $M=.60$ ,  $SD=.18$ ). Am seltensten wurden Items des wissenschaftlichen Niveaus gelöst (WV:  $M=.56$ ,  $SD=.15$ ; NV:  $M=.59$ ,  $SD=.22$ ). Diese Anordnung stellte sich so allerdings nur über alle Items hinweg dar. Auch bei Verdunstung wurden Items des mittleren Niveaus mit der höchsten Wahrscheinlichkeit gelöst ( $M=.65$ ,  $SD=.19$ ), die Mittelwerte für naive und wissenschaftliche Vorstellungen waren allerdings identisch ( $M=.60$ ). Bei Kondensation wurden naive Items am häufigsten gelöst ( $M=.59$ ,  $SD=.24$ ), während es keinen Unterschied der Lösungshäufigkeit auf dem mittleren und wissenschaftlichen Niveau gab ( $M=.53$ ). Insgesamt sind Aufgaben der Komponente Verdunstung leichter zu lösen gewesen ( $M=.61$ ,  $SD=.18$ ) als Items der Komponente Kondensation ( $M=.55$ ,  $SD=.19$ ).

Hinsichtlich der Kontextmerkmale zeigte sich, dass Aufgaben mit kalter Temperatur ( $M=.61$ ,  $SD=.18$ ) häufiger gelöst wurden als Aufgaben mit warmer Temperatur ( $M=.53$ ,  $SD=.18$ ). Allerdings war dieser Unterschied bei Verdunstung stärker ausgeprägt als bei Kondensation (siehe Tabelle 6). Desweiteren zeigte sich, dass Aufgaben, die innen situiert waren ( $M=.61$ ,  $SD=.18$ ), häufiger als Aufgaben draußen ( $M=.52$ ,  $SD=.17$ ) gelöst wurden. Diese Tendenz zeichnete sich auch für die beiden Konzepte separat ab (siehe Tabelle 6).

Das Merkmal sichtbarer Wasserspiegel wurde nur bei den Verdunstungsaufgaben betrachtet. Hier zeigte sich, dass Aufgaben, in denen der Wasserspiegel nicht zu sehen war, häufiger gelöst wurden (siehe Tabelle 6).

Ein Unterschied der Lösungshäufigkeit zeigte sich bei Kondensations-Items in Bezug darauf ob Wasser als sichtbarer Nebel in der Luft ( $M=.40$ ,  $SD=.05$ ) oder auf einer Fläche kondensierte ( $M=.56$ ,  $SD=.19$ ): Aufgaben, in denen Nebel entstand, wurden seltener gelöst. Darüber hinaus war

die mittlere Lösungshäufigkeit für Items, in denen keine potentielle Wasserquelle vorhanden ( $M=.46$ ,  $SD=.14$ ) war, geringer als in Items mit Wasserquelle ( $M=.58$ ,  $SD=.19$ ).

#### *4.4.2.2 Schwimmen und Sinken*

Auch für diesen Themenbereich sind in Tabelle 8 die mittleren Lösungshäufigkeiten und die zugehörigen Standardabweichungen für die unterschiedlichen Aufgabenkategorien dargestellt.

Über alle Komponenten hinweg wurden hier die MS-Aufgaben ( $M=.64$ ,  $SD=.15$ ) am häufigsten gelöst, gefolgt von MC- ( $M=.62$ ,  $SD=.62$ ) und FC-Aufgaben ( $M=.60$ ,  $SD=.13$ ). Deutlicher seltener gelöst wurden hingegen die Aufgaben mit offenem ( $M=.39$ ,  $SD=.05$ ) oder einem anderen Antwortformat ( $M=.33$ ,  $SD=.05$ ).

Ebenfalls sind bei diesem Themenbereich die Aufgaben mit einer Demonstration mit realen Gegenständen ( $M=.63$ ,  $SD=.17$ ) häufiger gelöst worden als solche ohne ( $M=.59$ ,  $SD=.14$ ). Dieser deskriptive Unterschied war allerdings nur sehr gering und bei Aufgaben zu Verdrängung nicht zu finden (siehe Tabelle 8).

In Bezug auf die Niveaus der Items zeigte sich ein leichter Vorsprung der naiven Vorstellungen ( $M=.64$ ,  $SD=.20$ ), während die Aufgaben des wissenschaftlichen Niveaus am seltensten gelöst wurden ( $M=.60$ ,  $SD=.14$ ). Diese Anordnung war allerdings für Aufgaben zum Auftrieb nicht wieder zu finden: Hier sind Zwischenvorstellungen am seltensten gelöst worden (siehe Tabelle 8).

Auch die Aufgaben, die den verschiedenen Ausprägungen der Kontextmerkmale zugeordnet wurden, unterscheiden sich deskriptiv in den relativen Lösungshäufigkeiten: Beim Auftrieb sind die Aufgaben seltener gelöst worden, in denen es um einen großen ( $M=.59$ ,  $SD=.16$ ), schweren ( $M=.57$ ,  $SD=.19$ ) oder hohlen ( $M=.60$ ,  $SD=.19$ ) Gegenstand ging, im Vergleich zu Aufgaben mit kleinen ( $M=.64$ ,  $SD=.15$ ), leichten ( $M=.64$ ,  $SD=.15$ ) oder ausgefüllten ( $M=.64$ ,  $SD=.14$ ) Gegenständen. Darüber hinaus waren die inkongruenten Dichte-Aufgaben ( $M=.48$ ,  $SD=.07$ ) am schwierigsten zu lösen, die kongruenten ( $M=.85$ ) am einfachsten. Bei Aufgaben zu Verdrängung waren jene leichter zu lösen, in denen nach einem Vergleich zwischen zwei Gegenständen gefragt wurde (Vergleich:  $M=.64$ ,  $SD=.15$ ; Einzelurteil:  $M=.43$ ,  $SD=.20$ ).

### **4.4.3 Ergebnisse der inferenzstatistischen Analysen von Studie 1**

#### *4.4.3.1 Verdunstung und Kondensation*

Die Prüfung der jeweiligen Haupteffekte der verschiedenen Aufgabenmerkmale auf die Aufgabenschwierigkeit mittels der generalisierten linearen gemischten Modelle wurde für die formalen Merkmalen, die Kompetenzniveaus und Basiskonzepte sowie für die Kontextmerkmale Ort und Temperatur über den gesamten Themenbereich hinweg durchgeführt (siehe Tabelle 10). Dabei zeigte sich nur ein marginal signifikanter Haupteffekt: Items zum Basiskonzept Kondensation waren etwas schwieriger zu lösen als Aufgaben zu Verdunstung ( $b=-0.39$ ,  $p<.10$ ), d.h. Kondensations-Items waren durchschnittlich  $-0.39$  logits schwieriger zu lösen als

Verdunstungs-Items. Die Varianzaufklärung, die durch Berücksichtigung dieses Effektes an den Itemschwierigkeiten erreicht wurde, war mit  $R^2=.04$  allerdings gering. Die im Vergleich höchste Varianzaufklärung zeigt sich in dem Modell, in dem der Effekt der Temperatur untersucht wurde ( $R^2=.07$ ). Obwohl dieses Modell zudem von den untersuchten Modellen die beste Passung auf die Daten aufzuweisen schien (BIC=41097), war der Effekt für das Aufgabenmerkmal allerdings nicht signifikant ( $b=-0.32$ ,  $p=.15$ ). Desweiteren ist Tabelle 10 zu entnehmen, dass keines der untersuchten Merkmale einen großen Beitrag zur Aufklärung der Schwierigkeitsvarianz leistete. Die Varianzaufklärung lag zwischen  $R^2=.01$  und  $R^2=.07$  und war damit in allen Fällen eher gering.

**Tabelle 10 Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Verdunstung und Kondensation**

	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	$R^2$
Nullmodell	0.39**	0.12	0.91	0.52	0.45	41796	
formale Aufgabenmerkmale							
Antwortformat			0.91	0.52	0.45	41817	.01
<i>(forced choice)</i> <sup>a</sup>	0.26	0.24					
<i>multiple select</i>	0.13	0.27					
<i>multiple choice</i>	0.44	0.47					
Demonstration mit realen Gegenständen			0.90	0.52	0.45	41804	.04
(nein)	0.30*	0.13					
ja	0.46	0.29					
Textlänge			0.91	0.52	0.45	41807	.00
(Intercept)	0.40	0.37					
Textlänge	0.00	0.00					
Merkmale aus Kompetenzmodell							
Niveaus			0.91	0.52	0.45	41817	.01
(NV)	0.41*	0.20					
ZV	0.12	0.30					
WV	-0.14	0.36					
Basiskonzept			0.89	0.52	0.45	41804	.04
(Intercept)	0.56***	0.16					
Kondensation	-0.39 <sup>+</sup>	0.22					
Kontextmerkmale							
Ort			0.89	0.52	0.45	41804	.04
(drinne)	0.49***	0.14					
draußen	-0.41	0.25					
Temperatur			0.88	0.52	0.45	41097	.07
(kalt)	0.49**	0.16					
warm	-0.32	0.24					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>+</sup> $p<.10$ ; \* $p<.05$ ; \*\* $p<.01$ ; \*\*\* $p<.001$ ; Die angegebene Varianzaufklärung ( $R^2$ ) bezieht sich auf den Anteil aufgeklärter Varianz am Residuum für Items. <sup>a</sup>Die Merkmalskategorie in Klammern wurde im jeweiligen Modell als Referenzkategorie genutzt. Das angegebene Regressionsgewicht (B) für diese Merkmalskategorie entspricht dem Y-Achsenabschnitt und kann als mittlere Schwierigkeit der Items der jeweiligen Kategorie interpretiert werden.

Auch bei Betrachtung der Effekte der Aufgabenmerkmale für Items zu Verdunstung zeigte sich lediglich für ein Kontextmerkmal ein marginal signifikanter Effekt (Tabelle 11): Items, die draußen

situierter waren, waren schwieriger zu lösen als Items, die drinnen situierter waren ( $b=-0.56$ ,  $p<0.10$ ). Dieses Merkmal wies eine Varianzaufklärung von 8% auf. Die höchste Varianzaufklärung ( $R^2=.13$ ) und die beste Modellpassung ( $BIC=20809$ ) waren auch bei Verdunstung für das Aufgabenmerkmal Temperatur zu finden und erneut nicht von einem signifikanten Haupteffekt dieses Merkmals begleitet.

**Tabelle 11 Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Verdunstung**

	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	R <sup>2</sup>
Nullmodell	0.57**	0.18	0.96	0.63	0.60	20987	
formale Aufgabenmerkmale							
Antwortformat			0.92	0.63	0.60	21004	.07
( <i>forced choice</i> ) <sup>a</sup>	0.99*	0.39					
<i>multiple select</i>	-0.58	0.42					
<i>multiple choice</i>	0.00	0.60					
Demonstration mit realen Gegenständen			0.93	0.63	0.60	209995	.05
(nein)	0.49**	0.18					
ja	0.71	0.50					
Textlänge			0.95	0.63	0.60	20997	.02
(Intercept)	0.99 <sup>+</sup>	0.53					
Textlänge	-0.01	0.01					
Merkmale aus Kompetenzmodell							
Niveaus			0.95	0.63	0.60	21006	.03
(NV)	0.49 <sup>+</sup>	0.27					
ZV	0.33	0.40					
WV	-0.02	0.38					
Kontextmerkmale							
Wasserspiegel			0.95	0.63	0.60	20997	.02
ja	0.36	0.30					
nein	0.30	0.35					
Ort			0.92	0.63	0.60	20994	.08
(drinnen)	0.76***	0.20					
draußen	-0.56 <sup>+</sup>	0.33					
Temperatur			0.89	0.63	0.60	20809	.13
(warm)	0.71***	0.21					
Sonne	-0.47	0.38					
kalt	-0.43	0.40					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>+</sup> $p<.10$ ; \* $p<.05$ ; \*\* $p<.01$ ; \*\*\* $p<.001$ ; Die angegebene Varianzaufklärung ( $R^2$ ) bezieht sich auf den Anteil aufgeklärter Varianz am Residuum für Items. <sup>a</sup>Die Merkmalskategorie in Klammern wurde im jeweiligen Modell als Referenzkategorie genutzt. Das angegebene Regressionsgewicht (B) für diese Merkmalskategorie entspricht dem Y-Achsenabschnitt und kann als mittlere Schwierigkeit der Items der jeweiligen Kategorie interpretiert werden.

Wie Tabelle 12 zu entnehmen ist, zeigten für die Items zu Kondensation zwei Aufgabenmerkmale einen signifikanten Einfluss auf die Aufgabenschwierigkeit: Items im Antwortformat *multiple select* waren leichter zu lösen als Aufgaben im Format *forced choice* ( $b=0.66$ ,  $p<.05$ ) und, da sich das weitere Aufgabenformat *multiple choice* nicht signifikant in der Schwierigkeit von der



Referenzkategorie unterscheidet ( $b=0.00$ ,  $p=.99$ ), lässt sich folgern, dass MS-Items zu Kondensation leichter zu lösen waren als Items mit anderen Formaten. Das Aufgabenmerkmal Antwortformat klärte 13% der Varianz auf Itemebene auf und konnte somit mehr zu Erklärung der Daten beitragen als die restlichen untersuchten Merkmale. Desweiteren waren Items, in denen keine potentielle Quelle des kondensierten Wassers genannt war, marginal signifikant leichter zu lösen als solche, in denen eine Quelle genannt war ( $b=0.63$ ,  $p<.10$ ). Das Modell mit diesem Aufgabenmerkmal wies eine Varianzaufklärung von 11% auf und zeigte außerdem von allen Modellen zur Komponente Kondensation die beste Passung auf die Daten ( $BIC=20323$ ).

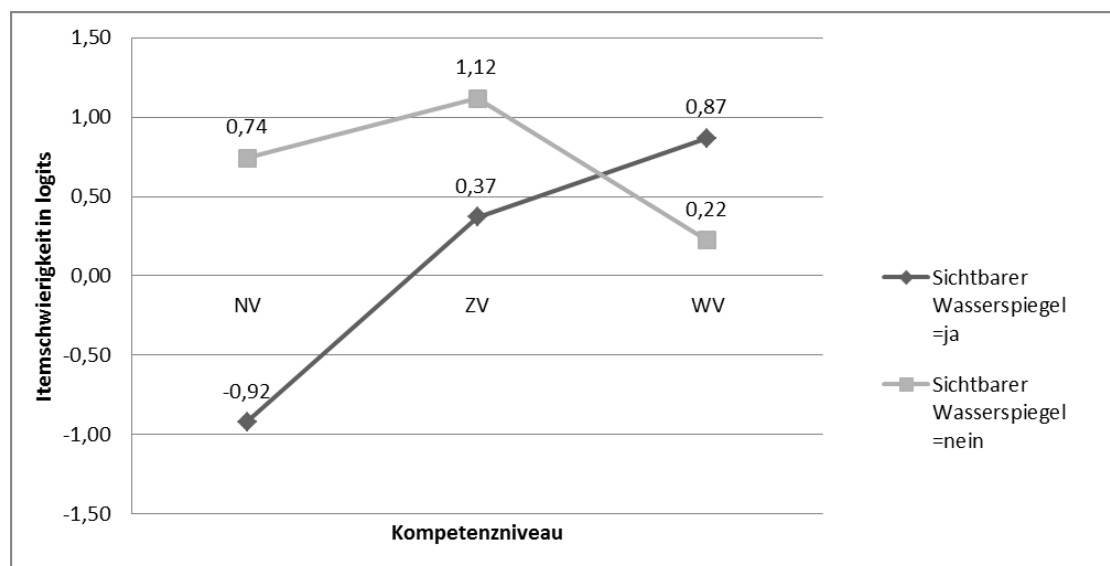
**Tabelle 12 Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Kondensation**

	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	R <sup>2</sup>
Nullmodell	0.21	0.16	0.85	0.49	0.42	20861	
formale Aufgabenmerkmale							
Antwortformat			0.80	0.49	0.42	20877	.13
<i>(forced choice)</i> <sup>a</sup>	-0.23	0.27					
<i>multiple select</i>	0.66*	0.32					
<i>multiple choice</i>	0.00	0.84					
Demonstration mit realen Gegenständen			0.82	0.49	0.42	20869	.07
(nein)	0.07	0.18					
ja	0.51	0.34					
Textlänge			0.84	0.49	0.42	20870	.04
(Intercept)	-0.29	0.51					
Textlänge	0.01	0.01					
Merkmale aus Kompetenzmodell							
Niveaus			0.85	0.49	0.42	20880	.02
(NV)	0.38	0.29					
ZV	-0.24	0.43					
WV	-0.23	0.37					
Kontextmerkmale							
Ort			0.83	0.49	0.42	20869	.05
(drinne)	0.29 <sup>+</sup>	0.18					
draußen	-0.50	0.41					
Temperatur			0.85	0.49	0.42	20326	.00
(warm)	0.23	0.22					
kalt	-0.11	0.32					
Kondensationskern			0.82	0.49	0.42	20324	.07
Fläche	0.26	0.17					
Dampf	-0.78	0.50					
Quelle			0.81	0.49	0.42	20323	.11
vorhanden	-0.27	0.29					
nicht vorhanden	0.63 <sup>+</sup>	0.34					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>+</sup> $p<.10$ ; \* $p<.05$ ; \*\* $p<.01$ ; \*\*\* $p<.001$ ; Die angegebene Varianzaufklärung (R<sup>2</sup>) bezieht sich auf den Anteil aufgeklärter Varianz am Residuum für Items. <sup>a</sup>Die Merkmalskategorie in Klammern wurde im jeweiligen Modell als Referenzkategorie genutzt. Das angegebene Regressionsgewicht (B) für diese Merkmalskategorie entspricht dem Y-Achsenabschnitt und kann als mittlere Schwierigkeit der Items der jeweiligen Kategorie interpretiert werden.

Neben den Haupteffekten wurden für jedes Aufgabenmerkmal Modelle berechnet in denen sowohl das Merkmal selbst sowie die Niveaus der Items als Prädiktoren eingeführt wurden. Durch den Vergleich von Modellen, in denen nur Haupteffekte untersucht wurden, mit solchen, in denen zusätzlich ein Interaktionsterm eingeführt war, mittels des  $\chi^2$ -Differenzentests konnte entschieden werden, ob die Interaktion zur Erklärung der Daten beitragen konnte. Die Ergebnisse aller gerechneten und verglichenen Modelle für diesen Teil der Ergebnisdarstellung sind in Anhang A05 zu finden. Für den vorliegenden Themenbereich zeigte sich lediglich für die Items zu Verdunstung, dass die Interaktion zwischen dem Merkmal Wasserspiegel und dem Kompetenzniveau einer Aufgabe einen signifikanten Beitrag zur Erklärung der Aufgabenschwierigkeit lieferte ( $\chi^2(2)=7.91$ ,  $p<.05$ ). Die Betrachtung der Regressionskoeffizienten dieses Modells offenbarte eine überproportional zunehmende Schwierigkeit der Items ohne sichtbaren Wasserspiegel mit steigendem Niveau (siehe Anhang A05.2). Die Abbildung 2 der durch das Modell geschätzten Mittelwerte der Itemschwierigkeiten offenbarte, dass Items des naiven und mittleren Niveaus ohne sichtbaren Wasserspiegel eine höhere Lösungshäufigkeit aufwiesen als solche mit. Dies kehrte sich für Items des wissenschaftlichen Niveaus um: Items mit sichtbarem Wasserspiegel wiesen hier eine höhere Lösungshäufigkeit auf. Neben dem signifikanten Erklärungsbeitrag der Interaktion am Gesamtmodell, war der Haupteffekt für das Kontextmerkmal Wasserspiegel in diesem Modell, unter Kontrolle der Niveaus, signifikant ( $b=1.66$ ,  $p<.05$ ): Items ohne sichtbaren Wasserspiegel waren durchschnittlich leichter zu lösen als Items mit. Auch der Unterschied zwischen naiven Items und mittleren Items ( $b=1.29$ ,  $p<.10$ ) und naiven und wissenschaftlichen Items ( $b=1.79$ ,  $p<.05$ ) war in diesem Modell signifikant. Die geschätzten Mittelwerte (siehe Abbildung 2) zeigten, dass der Schwierigkeitsunterschied zwischen Aufgaben mit und ohne sichtbaren Wasserspiegel auf dem naiven Niveau ( $\Delta=-1.66$ ) mehr als doppelt so groß war, wie auf dem wissenschaftlichen Niveau ( $\Delta=0.64$ ). Der Schwierigkeitsunterschied auf dem mittleren Niveau ( $\Delta=-0.75$ ) war ebenfalls deutlich kleiner als auf dem naiven Niveau, war im absoluten Betrag allerdings nicht wesentlich größer als auf dem höchsten Niveau.

**Abbildung 2 Mittlere geschätzte Schwierigkeiten der Items zu Verdunstung getrennt nach Kompetenzniveau und Sichtbarkeit des Wasserspiegels**



Anmerkungen. NV, naive Vorstellung; ZV, Zwischenvorstellung; WV, wissenschaftliche Vorstellung;

#### 4.4.3.2 Schwimmen und Sinken

Die Ergebnisse der Prüfung der Aufgabenmerkmale für den Themenbereich Schwimmen/Sinken sind in Tabelle 13 zu finden. Da keiner der untersuchten Kontextfaktoren auf die Items aller Komponenten (Auftrieb, Dichte, Verdrängung) anzuwenden war, wurden hier zunächst nur die formalen und Merkmale des Kompetenzmodells untersucht. Dabei zeigte sich, dass beim Antwortformat Aufgaben mit offenem oder graphischem Antwortformat signifikant schwieriger zu lösen waren als andere Aufgaben ( $b=-1.32$ ,  $p<.001$ ), wodurch eine Aufklärung der Schwierigkeitsvarianz um 22% erreicht werden konnte. Neben dem Antwortformat konnte keines der anderen Merkmale eine nennenswerte Varianzaufklärung erzielen.

**Tabelle 13 Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Schwimmen und Sinken**

	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	R <sup>2</sup>
Nullmodell	0.53***	0.12	0.83	0.79	0.56	48844	
formale Aufgabenmerkmale							
Antwortformat			0.73	0.79	0.56	48861	.22
( <i>forced choice</i> ) <sup>a</sup>	0.58*	0.27					
<i>multiple select</i>	0.09	0.29					
<i>multiple choice</i>	0.02	0.39					
offen/andere	-1.32**	0.42					
Demonstration mit realen Gegenständen			0.81	0.79	0.56	48853	.03
(nein)	0.34 <sup>+</sup>	0.19					
ja	0.28	0.22					
Textlänge			0.82	0.79	0.56	48854	.00
(Intercept)	0.63*	0.26					
Textlänge	-0.00	0.00					
Merkmale aus Kompetenzmodell							
Niveaus			0.82	0.79	0.56	48865	.00
(NV)	0.62**	0.21					
ZV	-0.12	0.33					
WV	-0.12	0.24					
Konzept			0.82	0.79	0.56	48864	.02
Verdrängung	0.31	0.26					
Dichte	0.29	0.36					
Auftrieb	0.29	0.24					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>+</sup> $p<.10$ ; \* $p<.05$ ; \*\* $p<.01$ ; \*\*\* $p<.001$ ; Die angegebene Varianzaufklärung (R<sup>2</sup>) bezieht sich auf den Anteil aufgeklärter Varianz am Residuum für Items. <sup>a</sup>Die Merkmalskategorie in Klammern wurde im jeweiligen Modell als Referenzkategorie genutzt. Das angegebene Regressionsgewicht (B) für diese Merkmalskategorie entspricht dem Y-Achsenabschnitt und kann als mittlere Schwierigkeit der Items der jeweiligen Kategorie interpretiert werden.

Für die Items zum Basiskonzept Auftrieb wiesen die formalen Merkmale Textlänge ( $b=0.02$ ,  $p<.05$ ; siehe Tabelle 14) und Demonstration mit realen Gegenständen ( $b=0.52$ ,  $p<.05$ ) einen signifikanten Effekt auf. Dabei erleichterten sowohl längere Texte wie auch die Demonstration mit

realen Gegenständen die Bearbeitung der Aufgaben. Diese beiden Modelle wiesen außerdem die höchsten Varianzaufklärungen für dieses Basiskonzept auf (Antwortformat:  $R^2=.15$ , Textlänge:  $R^2=.12$ ). Insgesamt ist für diesen Bereich festzustellen, dass alle formalen Aufgabenmerkmale eine höhere Erklärungskraft hatten ( $.10 < R^2 < .15$ ) als die Niveaus und die Kontextmerkmale ( $.01 < R^2 < .08$ ). Die beste Passung an die Daten wies anhand des gewählten Kriteriums das Modell mit dem Merkmal Größe auf (BIC=23122), welches unter den Kontextmerkmalen mit  $R^2=.08$  noch die meiste Varianz aufklärte, ohne allerdings einen signifikanten Haupteffekt dieses Merkmals aufzuweisen.

**Tabelle 14 Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Auftrieb**

	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	$R^2$
Nullmodell	0.56***	0.13	0.74	0.48	0.35	24460	
formale Aufgabenmerkmale							
Antwortformat			0.68	0.48	0.35	24473	.15
<i>(forced choice)</i> <sup>a</sup>	0.26	0.31					
<i>multiple select</i>	0.41	0.33					
offen	-0.78	0.58					
Demonstration mit realen Gegenständen			0.71	0.48	0.35	24466	.10
(nein)	0.17	0.23					
ja	0.52*	0.26					
Textlänge			0.70	0.48	0.35	24465	.12
(Intercept)	-0.20	0.35					
Textlänge	0.02*	0.00					
Merkmale aus Kompetenzmodell							
Niveaus			0.74	0.48	0.35	24479	.01
(NV)	0.62**	0.19					
ZV	-0.14	0.30					
WV	-0.06	0.28					
Kontextmerkmale							
Größe			0.71	0.51	0.35	23122	.08
(klein)	0.63***	0.14					
groß	-0.25	0.32					
Gewicht			0.73	0.48	0.35	24468	.04
(leicht)	0.64***	0.14					
schwer	-0.32	0.27					
Hohl			0.74	0.48	0.35	24429	.02
(gefüllt)	0.63***	0.15					
hohl	-0.19	0.25					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>†</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001; Die angegebene Varianzaufklärung ( $R^2$ ) bezieht sich auf den Anteil aufgeklärter Varianz am Residuum für Items. <sup>a</sup>Die Merkmalskategorie in Klammern wurde im jeweiligen Modell als Referenzkategorie genutzt. Das angegebene Regressionsgewicht (B) für diese Merkmalskategorie entspricht dem Y-Achsenabschnitt und kann als mittlere Schwierigkeit der Items der jeweiligen Kategorie interpretiert werden.

Items zum Bereich Dichte wurden, wie in Tabelle 15 dargestellt, in ihrer Lösungswahrscheinlichkeit signifikant durch die Variation des Merkmals Kongruenz beeinflusst.

Dabei waren sowohl Aufgaben in denen die Gegenstände das gleiche Gewicht aufwiesen ( $b=-2.02$ ,  $p<.001$ ) als auch inkongruente Aufgaben ( $b=-3.06$ ,  $p<.001$ ) schwerer zu lösen als kongruente Aufgaben und durch die Berücksichtigung dieses Faktors konnte eine große Varianzaufklärung von 87% erreicht werden.

**Tabelle 15 Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Dichte**

	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	R <sup>2</sup>
Nullmodell	0.73*	0.35	0.96	1.53	1.21	6986	
formale Aufgabenmerkmale							
Antwortformat			0.96	1.53	1.21	6995	.00
<i>(forced choice)</i> <sup>a</sup>	0.64	0.97					
<i>multiple choice</i>	0.10	1.01					
Kontextmerkmale							
Kongruenz			0.34	1.53	1.21	6987	.87
kongruent	2.90***	0.42					
gleiches Gewicht	-2.02***	0.44					
inkongruent	-3.06***	0.45					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; \* $p<.10$ ; \*\* $p<.05$ ; \*\*\* $p<.01$ ; \*\*\* $p<.001$ ; Die angegebene Varianzaufklärung (R<sup>2</sup>) bezieht sich auf den Anteil aufgeklärter Varianz am Residuum für Items. Da bei den weiteren Merkmalen in den Aufgaben zu Dichte keine Varianz vorlag, können diese nicht inferenzstatistisch auf ihre Relevanz für die Aufgabenschwierigkeit geprüft werden. <sup>a</sup>Die Merkmalskategorie in Klammern wurde im jeweiligen Modell als Referenzkategorie genutzt. Das angegebene Regressionsgewicht (B) für diese Merkmalskategorie entspricht dem Y-Achsenabschnitt und kann als mittlere Schwierigkeit der Items der jeweiligen Kategorie interpretiert werden.

Beim Basiskonzept Verdrängung hatten die Aufgabenmerkmale Antwortformat, Textlänge und Vergleich vs. Einzelurteil einen signifikanten Einfluss auf die Schwierigkeit (vgl. Tabelle 16): Items im Format *multiple choice* ( $b=-3.12$ ,  $p<.001$ ) waren schwieriger zu lösen als Items mit dem Format *forced choice* oder *multiple select*, wodurch ein großer Teil Varianz erklärt werden konnte ( $R^2=.59$ ). Auch Items mit längerem Text ( $b=-0.03$ ,  $p<.05$ ) waren zunehmend schwieriger zu lösen, was ebenfalls einen beachtlichen Anteil der Varianz erklärte ( $R^2=.20$ ). War in einer Aufgabe ein Vergleich zwischen Objekten gefragt, so war diese einfacher zu beantworten ( $b=1.53$ ,  $p<.05$ ) als Aufgaben, in denen Objekte separat beurteilt werden sollten. Dadurch konnten 24% der Varianz erklärt werden. In diesem Bereich wiesen die Modelle mit signifikanten Effekten auch die besten Passungen auf die Daten auf (Antwortformat: BIC=14349, Textlänge: BIC=14348, Vergleich vs. Einzelurteil: BIC=14348).

**Tabelle 16 Ergebnisse der konsekutiven Prüfung des Einflusses der Aufgabenmerkmale für Verdrängung**

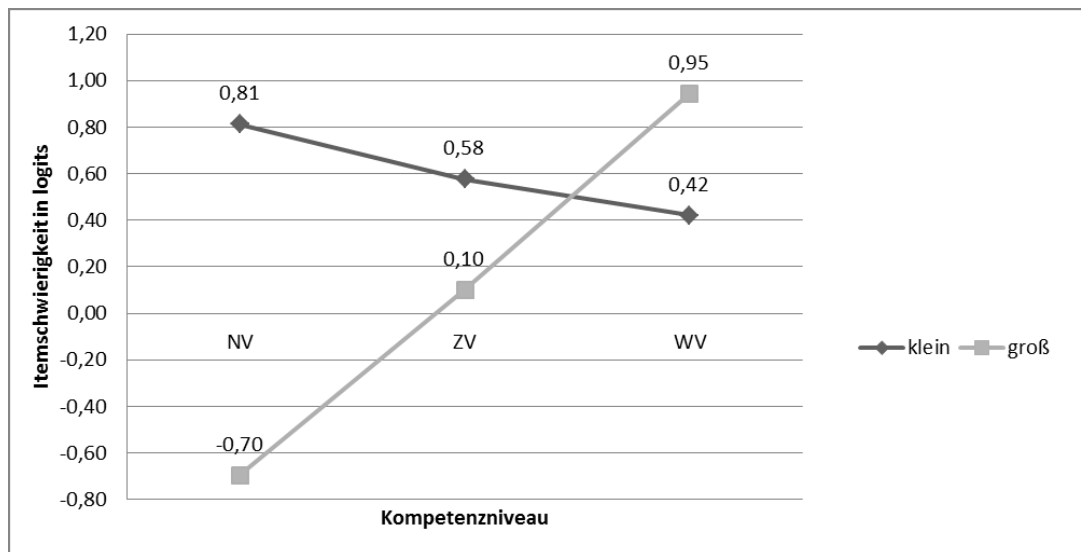
	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	R <sup>2</sup>
Nullmodell	0.51	0.43	1.44	2.41	1.39	14341	
formale Aufgabenmerkmale							
Antwortformat			0.92	2.41	1.39	14349	.59

<i>(forced choice)</i> <sup>a</sup>	1.76***	0.56					
<i>multiple select</i>	-0.99	0.64					
andere	-3.12***	0.76					
Demonstration mit realen Gegenständen			1.43	2.41	1.39	14351	.00
(nein)	0.35	1.45					
ja	0.17	1.50					
Textlänge			1.28	2.41	1.39	14348	.20
(Intercept)	3.00*	1.44					
Textlänge	-0.03 <sup>+</sup>	0.02					
<hr/>							
Niveaus			Merkmale aus Kompetenzmodell				
(NV)	0.43	1.03	1.44	2.41	1.39	14351	.00
WV	0.10	1.11					
<hr/>							
Vergleich			Kontextmerkmale				
Einzelurteil	-0.55	0.65	1.25	2.41	1.39	14348	.24
Vergleich	1.53*	0.75					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001; Die angegebene Varianzaufklärung (R<sup>2</sup>) bezieht sich auf den Anteil aufgeklärter Varianz am Residuum für Items. <sup>a</sup>Die Merkmalskategorie in Klammern wurde im jeweiligen Modell als Referenzkategorie genutzt. Das angegebene Regressionsgewicht (B) für diese Merkmalskategorie entspricht dem Y-Achsenabschnitt und kann als mittlere Schwierigkeit der Items der jeweiligen Kategorie interpretiert werden.

In dem Themenbereich Schwimmen/Sinken konnten nur für die Items zu Auftrieb die Interaktionen zwischen Kontextmerkmalen und Niveaus untersucht werden (siehe Anhang A06). Dabei stellte sich nur für die Interaktion zwischen dem Merkmal Größe und dem Niveau einer Aufgabe ein substantieller Erklärungsbeitrag zusätzlich zu den Haupteffekten heraus ( $\chi^2(2)=6.20$ ,  $p=.05$ ). Wie in Abbildung 3 zu erkennen ist, sind Items des naiven oder Zwischenniveaus leichter zu lösen, wenn es um einen kleinen Gegenstand geht. Wissenschaftliche Items allerdings wurden häufiger gelöst, wenn es um große Gegenstände ging. Die geschätzten Mittelwerte (siehe Abbildung 3) offenbaren für das mittlere ( $\Delta=0.48$ ) und das wissenschaftliche ( $\Delta=-0.53$ ) Niveau einen geringeren Unterschied in den Schwierigkeiten zwischen Aufgaben mit kleinen und großen Objekten als auf dem untersten Kompetenzniveau ( $\Delta=1.51$ ).

**Abbildung 3 Mittlere geschätzte Schwierigkeiten der Items zu Auftrieb getrennt nach Kompetenzniveau und Größe des Gegenstandes**



Anmerkungen. NV, naive Vorstellung; ZV, Zwischenvorstellung; WV, wissenschaftliche Vorstellung;

## 4.5 Diskussion von Studie 1

### 4.5.1 Zusammenfassung und Diskussion der Ergebnisse von Studie 1

Die Querschnittstudie verfolgte das Ziel die in Kapitel 3 aufgeworfenen Fragestellungen nach dem Wirken von Kontextmerkmalen an einer quasi natürlich angefallenen Itemstichprobe, die nicht speziell für die Untersuchung des Einflusses der Merkmale entwickelt worden war, zu untersuchen. Die Relevanz dieser Fragen wurde zuvor mit der schon per Definition gegebenen situationsbezogenen Natur von Schülerkompetenzen (vgl. Kapitel 2.1.1) und der sowohl theoretisch wie empirisch bedeutsamen Kontextabhängigkeit von Schülervorstellungen (vgl. Kapitel 2.2.2 und 2.2.3) begründet. Dazu wurden in Kapitel 2.3.3.2 und 2.3.3.3 Merkmale von Aufgabenkontexten in der existierenden Literatur zu Schülervorstellungen in den untersuchten Themenbereich identifiziert, die möglicherweise einen Einfluss auf die Aufgabenbearbeitung und damit die empirischen Aufgabenschwierigkeiten haben würden, und eine angenommene Wirkrichtung dieser Merkmale beschrieben. Im Folgenden werden die Ergebnisse aus Studie 1 zusammengefasst und hinsichtlich ihrer Bedeutung für die in Kapitel 4.2 aufgeworfenen Fragestellungen diskutiert.

#### 4.5.1.1 F1: Vorliegen der Kontextmerkmale in den Aufgaben von Studie 1

Frage 1 bezog sich auf das Vorliegen der interessierenden Merkmale in den Aufgaben zur Erfassung naturwissenschaftlicher Kompetenz in der Grundschule. Neben den zuvor bereits erwähnten Kontextmerkmalen wurden ausgewählte formale Aufgabenmerkmale sowie Merkmale im Zusammenhang mit dem der Aufgabenentwicklung zugrunde liegenden Kompetenzmodell, d.h.

Kompetenzniveau und Komponente, in den Blick genommen. Die in Kapitel 4.4.1 dargestellten Ergebnisse zeigten, dass alle Ausprägungen der Kontextmerkmale in den Aufgaben der Untersuchung vertreten waren. Bei genauerer Betrachtung zeigte sich allerdings, dass nicht alle Ausprägungen für jede getestete Komponente realisiert wurden.

So fiel insbesondere beim Themenbereich Schwimmen/Sinken auf, dass bei keiner der drei Komponenten Auftrieb, Dichte und Verdrängung alle Aufgabenformate realisiert wurden und dass bei Dichte in keinem Fall eine Demonstration mit realen Gegenständen möglich war. Darüber hinaus wurde das Niveau der Zwischenvorstellungen für Dichte und Verdrängung nicht angenommen und das naive Niveau wurde lediglich für Verdrängung in zwei Items untersucht. Die Unregelmäßigkeiten im Hinblick auf die Niveaus ergaben sich, wie in Kapitel 4.3.1.1 beschrieben, aus sorgfältigen methodischen und inhaltlichen Abwägungen, die im Verlaufe der Instrumententwicklung im Projekt Science-P aufgetreten waren. Diese Konfundierung ist also als Spezifikum des jeweiligen Inhalts anzusehen: Während es im Themenbereich Verdunstung/Kondensation und auch bei der Komponente Auftrieb jeweils mehrere Aspekte gibt, die verstanden werden müssen, wie z.B. der Massenerhalt, der Wechsel des Aggregatzustandes und Ortes für Verdunstung oder die Bedeutung der Auftriebskraft für Schwimmen bzw. Sinken und das Wechselspiel von Auftriebs- und Gewichtskraft, beziehen sich die Aufgaben zu Dichte und Verdrängung auf das Verständnis eines enger umgrenzten Aspektes. Bei den Aufgaben zu Dichte geht es um den Vergleich der relativen Dichte von je zwei Objekten. Das Verständnis der Abhängigkeit der verdrängten Wassermenge vom Volumen eines Gegenstandes und nicht von der Masse wird in Verdrängungsaufgaben getestet. Die Erlangung dieser beiden Kompetenzen führt über einen direkteren Weg als bei den anderen Aspekten zu einer wissenschaftsadäquaten Vorstellung, bei dem nur selten Vorstellungen des mittleren Niveaus auftreten. Zugleich ist die Anzahl der bekannten und häufig genannten naiven Vorstellungen für diese Aspekte geringer. Diese inhaltlichen Besonderheiten führen zudem dazu, dass die Konstruktion plausibler Antwortalternativen, wie sie für MS-Aufgaben verwendet werden, schwieriger ist. Somit liegt auch die Konfundierung im Hinblick auf Antwortformate im spezifischen Inhalt der Items begründet.

In Bezug auf die Merkmale des Kontexts zeigte sich bei beiden Themenbereichen, dass zwar alle Ausprägungen in den Aufgaben vertreten waren, häufig aber unterschied sich die Anzahl der zugeordneten Items deutlich. Beispielsweise waren im Themenbereich Verdunstung/Kondensation sowohl draußen wie auch drinnen situierte Items zu finden. Eine Beschreibung von Situationen innerhalb geschlossener Räume war allerdings fast dreimal so häufig wie außerhalb. Besonders augenfällig war dies in diesem Themenbereich für die Sichtbarkeit des Wasserspiegels, den Ort, den Kondensationskern und das Vorhandensein einer potentiellen Wasserquelle. Bei Schwimmen/Sinken zeigte sich dies bei allen Kontextmerkmalen. Diese Ungleichverteilung kann bei inferenzstatistischen Auswertungen zu höheren Standardfehlern führen, da Schätzungen, die auf weniger Fällen beruhen, nur mit eingeschränkter Präzision ermittelt werden können. Das



uneinheitliche Vorliegen der Merkmale kann mehrere Ursachen haben: Zunächst wurden während der Itementwicklung die Antwortformate und Kompetenzniveaus als relevante Faktoren berücksichtigt, während die Auswahl von Aufgabenkontexten sich subjektiv an der Erfahrungswelt von Grundschulern sowie als Lernkontext genutzten Situationen orientierte. Die Kontextmerkmale wurden dabei nicht explizit als Faktoren in Erwägung gezogen. Die häufiger vertretenen Ausprägungen könnten tatsächlich eine höhere Prävalenz in der Lebenswelt von Grundschulern haben, dies könnte allerdings nur durch weitergehende Studien eruiert werden.

Über das Vorliegen der Aufgabenmerkmale hinaus wurden mögliche Zusammenhänge zwischen den Merkmalen untersucht. Dies bezog sich auf Beziehungen zwischen den Kontextmerkmalen und dem Kompetenzniveau eines Items sowie auf die Multikollinearität (siehe z.B. Bortz, 1999) der Kontextmerkmale untereinander. Dabei zeigte keines der situativen Merkmale einen Zusammenhang zu den Niveaus, allerdings fanden sich viele hohe Korrelationen der situativen Merkmale untereinander. Im Themenbereich Schwimmen/Sinken zeigte sich insbesondere ein starker Zusammenhang zwischen der Größe und dem Gewicht der beschriebenen Objekte für die Auftrieb-Items. Dies ist wohl darauf zurückzuführen, dass in der physikalischen Umwelt tatsächlich häufig größere Gegenstände auch schwerer sind, vor allem in der subjektiven Wahrnehmung. Ausnahmen sind nur bei Gegenständen aus Materialien mit sehr geringer (z.B. Styropor) oder sehr hoher (z.B. Blei) Dichte zu finden. Der signifikante Zusammenhang zwischen dem Gewicht und der Hohlheit von Objekten erscheint zunächst kontra-intuitiv, da erwartungsgemäß nicht ausgefüllte Gegenstände leichter sein sollten. Dieser Zusammenhang ist allerdings wiederum auf die unausgewogene Verteilung der Items auf Merkmalsausprägungen zurück zu führen, da insgesamt nur wenige Items mit den Eigenschaften schwer oder hohl zu finden sind.

Da in innen situierten Aufgaben ein angenommener Einfluss der Sonne nicht möglich ist, war ein signifikanter Zusammenhang zwischen den Kontextmerkmalen Ort und Temperatur im Themenbereich Verdunstung/Kondensation zu finden. Für die Komponente Kondensation war dieser Zusammenhang nicht zu finden, da der Einfluss der Sonne in diesen Items unberücksichtigt blieb. Betrachtet man die Komponente Verdunstung separat, zeigten sich komplexe Zusammenhänge zwischen allen betrachteten situativen Merkmalen: Die meisten Aufgaben mit sichtbarem Wasserspiegel waren drinnen situiert. Daraus ergibt sich, dass ebenfalls bei den meisten Aufgaben mit sichtbarem Wasserspiegel nicht ein Einfluss der Sonne angenommen werden kann. Ebenso war bei Items zu Kondensation eine starke Beziehung zwischen den Merkmalen Kondensationskern, Quelle, Ort und Temperatur zu finden. Diese komplexe Multikollinearität der Kontextfaktoren in diesem Themenbereich ist u.a. dadurch begründet, dass die Prozesse des Verdunstens und Kondensierens bestimmte physikalische Voraussetzungen erfordern. Beispielsweise kondensiert Wasser dann, wenn die Gasphase, z.B. die Raumluft, mit Wasserdampf

übersättigt ist. Diese Bedingung wiederum hängt von weiteren Umweltfaktoren wie der Lufttemperatur ab.

Insgesamt war durch das Vorliegen der Multikollinearität die Möglichkeit zur Untersuchung von Interaktionen zwischen Merkmalen bzw. die Vorhersage der Itemschwierigkeiten durch mehrere Merkmale eingeschränkt, da durch sie die Schätzung von Regressionskoeffizienten instabil und ungenau wird und die Interpretation der resultierenden Koeffizienten erschwert ist (vgl. z.B. Morrison, 2003).

#### *4.5.1.2 F2: Effekt der Kontextmerkmale auf die Aufgabenschwierigkeiten in Studie 1*

Die nächste Fragestellung der Arbeit (Frage 2) fragte nach dem Einfluss der Kontextmerkmale auf die empirische Itemschwierigkeit. Insgesamt zeigten nur wenige der untersuchten Merkmale einen signifikanten Effekt, von diesen gefundenen Effekten waren einige nur marginal signifikant. Im ersten Themenbereich zeigte sich für die Items zu Verdunstung ein marginal signifikanter Effekt des Ortes, d.h. draußen situierte Aufgaben waren etwas schwieriger zu lösen als solche innerhalb geschlossener Räumlichkeiten. In Situationen außerhalb geschlossener Räumlichkeiten ist die Assoziation mit dem Wasserkreislauf als ganzem salienter. Einflussfaktoren wie die Sonne oder Wolken, die häufig in naiven Vorstellungen von Schülern eine Rolle spielen (Osborne & Cosgrove, 1983; Russell et al., 1989; Tytler, 2000), können auch in diesen Situationen zur Erklärung heran gezogen werden, während innerhalb geschlossener Räume der Einfluss dieser Größen nicht plausibel ist. Allerdings war der entsprechende Unterschied für Aufgaben zu Kondensation nicht zu finden. Obwohl anzunehmen ist, dass die gleichen Erschwernisse auch für Kondensationsaufgaben eine Rolle spielen könnten, ist zu beachten, dass die Kondensationsaufgaben insgesamt einen höheren Schwierigkeitsgrad aufwiesen als Aufgaben zu Verdunstung. Da die Probanden der Studie also insgesamt größere Schwierigkeiten beim Verständnis dieses Aspektes hatten, sind auch die innen situierten Aufgaben zu Kondensation schwieriger für sie. Zudem ist die Verteilung der Items auf die Merkmale bei Kondensation deutlich extremer: nur ein Sechstel der Aufgaben ist draußen situiert. Die Schätzung des Regressionsgewichtes fällt daher ungenauer aus.

Die Sichtbarkeit des Wasserspiegels zeigte in dieser Studie keinen signifikanten Haupteffekt auf die Itemschwierigkeiten, obwohl frühere Studien eine gesteigerte Prävalenz von naiven Vorstellungen, die das Einziehen von Wasser thematisieren, für Aufgaben ohne sichtbaren Wasserspiegel berichteten (Bar & Galili, 1994; Osborne & Cosgrove, 1983; Russell et al., 1989). In dieser Arbeit wurde allerdings im Vergleich zu früheren Studien ein anderes Antwortformat verwendet: Der Einsatz eines Fragebogens mit geschlossenem Antwortformat, im Vergleich zu offenen Interviews (z.B. Osborne & Cosgrove, 1983), führt häufig zu einem höheren Niveau der Aufgabenlösung (vgl. Pollmeier et al., 2011). Es ist daher möglich, dass die Schüler ihre Vorstellung, dass Wasser in die Materialien einzieht, in dieser Studie aufgrund des

Antwortformates in allen Fällen leichter überwinden konnten. Auch der Faktor Temperatur beeinflusste das Antwortverhalten auf Verdunstungsitems in dieser Studie nicht signifikant.

Items zu Kondensation, in denen keine potentielle Quelle des kondensierten Wasser vorhanden war, waren signifikant leichter zu lösen als Items, in denen eine Quelle vorhanden war. Das Verständnis von Kondensation fällt insbesondere jüngeren Schülern oftmals deswegen schwer, weil sie noch keine Vorstellung von dem gasförmigen Zustand haben und sich somit nicht vorstellen können, dass ständig gasförmiges Wasser in der Luft ist, welches unter entsprechenden Bedingungen kondensiert. Ist nun sichtbares, flüssiges Wasser in einer Situation vorhanden, kann dieses plausibel für Erklärungen genutzt werden. Ist dies allerdings nicht der Fall, muss etwas anderes die Quelle des kondensierten Wassers sein. Die Luft in diesem Fall mit in die Überlegungen einzubeziehen liegt näher und die Bearbeitung der Aufgaben wird so evtl. erleichtert. Die Kontextmerkmale Ort, Temperatur und Kondensationskern beeinflussten die Schwierigkeit der Items zu Kondensation nicht signifikant. V.a. in Bezug auf das Merkmal Kondensationskern führte die deutliche Ungleichverteilung der Items allerdings zu großen Standardfehlern der Schätzungen.

Bei den Ergebnissen im Themenbereich Schwimmen/Sinken ist auffällig, dass für die Komponenten Dichte und Verdrängung die untersuchten Kontextmerkmale zur Erklärung der Schwierigkeitsvarianz beitragen konnten, aber für die Items zu Auftrieb kein erklärungs mächtiger situativer Faktor gefunden wurde. Auch die deskriptiven Unterschiede in den mittleren Lösungshäufigkeiten waren bei den Auftriebits items geringer als bei den anderen Items. Obwohl die untersuchten Kontextmerkmale häufig zentrale Aspekte in naiven Vorstellungen zu diesem Bereich sind, wurde das Antwortverhalten nicht durch diese Merkmale beeinflusst. Allerdings wurden in diesem Bereich auch mehrere verschiedene Aspekte der Auftriebskraft im Zusammenhang mit dem Schwimmverhalten erfragt. Möglicherweise waren die untersuchten Merkmale nicht für alle dieser Aspekte relevant oder beeinflussten das Antwortverhalten bzgl. unterschiedlicher Aspekte in verschiedene Richtungen. Darüber hinaus sind hier auch die naiven und mittleren Vorstellungen vielfältiger als in den anderen untersuchten Themenbereichen und Schüler halten häufiger mehrere Vorstellungen parallel (vgl. Pollmeier et al., 2011). Systematische Unterschiede in diesen komplexen und diversen Gedankengebäuden zu finden, wird durch dieses Spezifikum des Inhaltsgebietes erschwert.

Ein besonders erklärungs mächtiges Merkmal war die Kongruenz der Items zum Bereich Dichte. Kongruente Aufgaben, bei denen das schwerere Objekt auch die höhere Materialdichte aufwies, waren deutlich leichter zu lösen als inkongruente Aufgaben, bei denen das schwerere Objekt die geringere Dichte hatte. Ebenso waren auch Aufgaben, in denen die Größe Gewicht über die Objekte konstant gehalten war, schwerer zu lösen als kongruente Aufgaben. Der Schwierigkeitsunterschied zwischen kongruenten und inkongruenten Dichteaufgaben deckt sich mit Befunden von Fassouloupoulos und Kollegen (2003) und Smith und Kollegen (1985). Während

bei den kongruenten Aufgaben noch keine vollständige Differenzierung der Begriffe Dichte und Gewicht erfolgt sein muss, um diese zu lösen, ist diese bei inkongruenten Aufgaben eine Voraussetzung für eine korrekte Lösung. Ebenso bietet bei Aufgaben, in denen beide Objekte das gleiche Gewicht haben, dies keinen Hinweis auf die richtige Antwort und die Differenzierung der beiden Konzepte ist notwendig.

Auch bei den Items zu Verdrängung hatte das untersuchte Kontextmerkmal einen signifikanten Einfluss auf die Aufgabenschwierigkeit. Aufgaben, in denen die Verdrängung durch verschiedene Objekte verglichen werden sollte, waren einfacher zu lösen als Aufgaben, in denen die Verdrängung durch einen Gegenstand erklärt werden sollte. Durch den Vergleich zweier Gegenstände, die sich hinsichtlich ihres Gewichts und ihres Volumens unterschieden, scheint den Schülern die Überwindung ihrer Vorstellung, die Verdrängung hänge vom Gewicht ab, zugunsten der angemesseneren Vorstellung, die Verdrängung hängt vom Volumen ab, erleichtert worden zu sein. Bei der Betrachtung eines einzelnen Objektes und der Erklärung der Verdrängung scheint für sie das Gewicht die salientere Größe gewesen zu sein.

Insgesamt zeigte sich für den Themenbereich Verdunstung/Kondensation eine stärkere Kontextualisierung des Antwortverhaltens, die durch die hier untersuchten Merkmale erklärt werden konnte, d.h. mehrere der untersuchten Merkmale wiesen einen Effekt auf, der allerdings häufig von eher geringer Stärke war. Im Themenbereich Schwimmen/Sinken hingegen konnten für zwei der untersuchten Komponenten zwar besonders erklärungsstarke Faktoren identifiziert werden, für die Komponente Auftrieb hingegen konnte kein systematischer Einfluss der Kontextmerkmale auf die Schwierigkeit festgestellt werden. Dieser allgemeine Unterschied zwischen den Themenbereichen deckt sich mit den Befunden zu Schülervorstellungen in den jeweiligen Bereichen, die in Kapitel 2.2.4 und 2.2.5 zusammen gefasst sind. In den Studien zu Verdunstung und Kondensation berichteten die Autoren (z.B. Costu & Ayas, 2005; Russell et al., 1989; Tytler, 2000) von konkreten Unterschieden in den geäußerten Schülervorstellungen, die auf bestimmte Aufgabenkontexte zurückgeführt werden konnten. Im Themenbereich Schwimmen/Sinken berichteten die Autoren eher von individuellen Erklärungen für unterschiedliche Objekte ohne allerdings eine Systematik erkennen zu lassen (z.B. Biddulph & Osborne, 1984). In Bezug auf die Kontextmerkmale, die keinen signifikanten Einfluss zeigten, ist allerdings anhand der berichteten Ergebnisse nicht zu erkennen, ob die Merkmale tatsächlich keinen Einfluss auf das Antwortverhalten haben, oder ob ein möglicher Einfluss nicht für alle Personen in die gleiche Richtung wirkt. Ein möglicher differentieller Einfluss könnte sich dabei für Probandengruppen oder auch für Individuen zeigen.

#### *4.5.1.3 F3: Interaktion zwischen Kontextmerkmalen und Kompetenzniveaus*

Eine mögliche differentielle Wirkung der Kontextmerkmale könnte sich auch in Bezug auf die unterschiedlichen Kompetenzniveaus zeigen. In Frage 3 sollte daher eine mögliche Interaktion

zwischen Kontextmerkmalen und Kompetenzniveaus untersucht werden. In den hierarchischen Modellvergleichen, in denen jeweils ein Modell mit den Haupteffekten der Kompetenzniveaus und des jeweiligen Kontextmerkmals mit einem Modell mit zusätzlichem Interaktionsterm verglichen wurde, konnte die Interaktion nur in zwei Fällen zu einer signifikant verbesserten Erklärung der Daten beitragen. Für die Items zu Verdunstung zeigte das Modell mit der Interaktion zwischen dem Kontextmerkmal Wasserspiegel und dem postulierten Niveau eines Items eine bessere Passung auf die Daten. Die Interaktion zwischen der Größe des in einer Aufgabe betrachteten Objektes und dem Niveau wirkte sich signifikant auf der Erklärung der Items zur Komponente Auftrieb aus. In beiden Fällen war der Schwierigkeitsunterschied zwischen Aufgaben unterschiedlicher Merkmalsausprägungen auf dem naiven Niveau deutlicher ausgeprägt als auf dem mittleren oder wissenschaftlichen Niveau, d.h. fachlich angemessenere Vorstellung waren in diesen beiden Fällen weniger stark von einer Kontextualisierung betroffen. Beide Merkmale zeigten keinen signifikanten Einfluss auf die Aufgabenschwierigkeit in den Modellen, in denen die Merkmale separat betrachtet wurden, hatten aber jeweils ein signifikantes Regressionsgewicht in den Modellen mit Interaktion. Die Kontextabhängigkeit der Vorstellungen war in diesen Fällen also nicht universell zu beobachten, sondern zeigte sich erst in einer differenzierteren Betrachtungsweise der Aufgaben. In den Fällen, in denen in dieser Studie eine Kontextualisierung von Vorstellungen durch Haupteffekte der jeweiligen Merkmale belegt werden konnte, zeigte sich allerdings keine signifikante Interaktion zu den Kompetenzniveaus, d.h. keine zunehmende Dekontextualisierung mit steigender fachlicher Angemessenheit der Vorstellungen. Möglicherweise werden die unterschiedlichen Merkmalsausprägungen in diesen Fällen von den Probanden als so bedeutsam wahrgenommen, dass Vorstellungen aller Niveaus nach Kontexten differenziert angewendet werden. Eine weitere mögliche Erklärung könnte die Sophistizierung des Wissens der Probanden allgemein sein: Evtl. wäre durch weiteren Unterricht in diesen Bereichen auch eine Unabhängigkeit der Erklärungen zu erreichen.

#### *4.5.1.4 F4: Effekt der formalen Merkmale auf die Aufgabenschwierigkeiten in Studie 1*

Die letzte Fragestellung (Frage 4) zu Studie 1 behandelte die Erklärungsmacht der untersuchten situativen Attribute. Dazu wurde zunächst der Einfluss von formalen Aufgabencharakteristiken, die sich in früheren Studien als relevant herausstellten (z.B. Pollmeier et al., 2011; Prenzel et al., 2002), auf die empirischen Itemschwierigkeiten untersucht. Die gefundenen Ergebnisse konnten dann in ihrer Stärke mit den Effekten der situativen Merkmale verglichen werden. Da die Kontextmerkmale in engerem Zusammenhang mit dem inhaltlich untersuchten Konstrukt stehen, sollten diese eine größere Erklärungskraft aufweisen. Zunächst ist darauf hinzuweisen, dass die erzielte Varianzaufklärung in den Modellen für einen gesamten Themenbereich (z.B. Verdunstung/Kondensation) jeweils geringer ausfiel als die Varianzaufklärung, die durch die Modelle für die jeweiligen Komponenten (z.B. Verdunstung) erzielt werden konnte. Im Bereich

Verdunstung konnten Modelle, die den Einfluss von Kontextmerkmalen untersuchten, am meisten Variabilität erklären. Bei Kondensationsitems klärte das Modell mit dem Merkmal Aufgabenformat die meiste Varianz auf, allerdings konnten auch Modelle mit Kontextmerkmalen einen Beitrag zur Aufklärung leisten. Im Gegensatz dazu leisteten im Themenbereich Schwimmen/Sinken v.a. im Bereich Auftrieb die formalen Merkmale eine höhere Varianzaufklärung. Für Items zu Verdrängung besaß zwar auch das Kontextmerkmal eine große Erklärungsmacht. Dennoch klärte das Aufgabenformat mehr Varianz auf. Da in den beiden Bereichen, in denen das Antwortformat Schwierigkeitsvarianz erklären konnte, die Offenheit des Aufgabenformates variiert wurde, welches in den anderen Bereichen stets geschlossen war, zeigt dies deutlich, dass im Bereich der Grundschule die Aufgabenanforderung durch Variation des Antwortformates stark beeinflusst werden kann (vgl. auch Bullock, Sodian & Koerber, 2009; Pollmeier et al., 2011). Anders stellen sich die Befunde im Bereich Dichte dar: Hier konnte der Großteil der Schwierigkeitsvarianz durch das situative Aufgabenmerkmal aufgeklärt werden. Berücksichtigt man nun zur Beurteilung dieser Frage zudem noch die Merkmale, die im Zusammenhang mit den Kompetenzmodell untersucht wurden (d.h. Kompetenzniveaus und Komponenten), ist festzuhalten, dass durch diese Merkmale nur eine geringe Erklärung der Itemstreuung erreicht werden konnte. Insgesamt zeigten die Kontextmerkmale keine deutlich höhere Erklärungskraft als die formalen Merkmale, bezogen auf einzelne Bereiche allerdings trugen sie dazu bei, Schwierigkeitsvarianz zu erzeugen.

#### *4.5.1.5 Zusammenfassung*

Zusammenfassend ist in dieser Studie kein Hinweis auf eine starke Kontextualisierung des Wissens zu finden, die für alle Personen gleich funktioniert bzw. sich an den untersuchten Kontextmerkmalen ausrichtet. Bei differenzierter Betrachtung der beiden untersuchten Themenbereiche zeigt sich für Verdunstung/Kondensation eine stärkere Ausrichtung der Itemschwierigkeiten an situativen Merkmalen als der Aufgaben als für Schwimmen/Sinken. Dieser Unterschied zwischen den Themenbereichen ist vermutlich auf inhaltliche Spezifika der Schülervorstellungen in den Themenbereich zurück zu führen: Während bei Verdunstung und Kondensation naive Vorstellungen weniger vielfältig sind (vgl. Pollmeier et al., 2011) und sich auch in früheren Studien bereits systematisch Diskrepanzen im Hinblick auf unterschiedliche Aufgabenkontexte feststellen ließen (z.B. Tytler, 2000), finden sich zu Schwimmen/Sinken mehr verschiedenartige Vorstellungen und Autoren berichteten darüber hinaus, dass individuelle Erklärungen für einzelne Objekte konstruiert wurden (z.B. Biddulph & Osborne, 1984). Auch aus fachlicher Sicht ist diese Diskrepanz zwischen den Themenbereichen wieder zu entdecken. Während die Studie für den Themenbereich Verdunstung/Kondensation auf drei klar umrissene Aspekte bzw. deren Verständnis fokussierte (Massenerhalt, Veränderung des Aggregatzustandes, Wechsel des Ortes), behandelten die Aufgaben zu Schwimmen und Sinken, speziell zur Komponente Auftrieb, mehrere Aspekte (Auftriebskraft und ihre Rolle für das Schwimmverhalten,

Wechselspiel zwischen Auftriebs- und Gewichtskraft, Bedeutung der relativen Dichte für das Schwimmverhalten, proportionales Dichteverständnis, Abhängigkeit der Verdrängung vom Volumen des Objektes). Diese Ergebnisse weisen nur tendenziell auf ein anfängliches Wissen, welches sich an situativen Merkmalen der Aufgaben orientiert.

#### **4.5.2 Beschränkungen der Studie 1**

Methodische Beschränkungen der vorliegenden Studie finden sich auf der Ebene der Aufgabenkonstruktion, des Testheftdesigns und der Stichprobe bzw. der Verteilung der Stichprobe auf die Testhefte. Darüber hinaus sind einige allgemeine Anmerkungen zu machen.

Die Itementwicklung im Projekt Science-P orientierte sich an der von Wilson (2005) vorgeschlagenen Methode der konstruktbasierten Instrumententwicklung. Dabei wurden zentrale inhaltliche Merkmale der untersuchten Kompetenz vorab theoretisch präzisiert und waren leitende Prinzipien bei der Aufgabenkonstruktion. Die in meiner Arbeit untersuchten Kontextmerkmale waren nicht Teil dieser Operationalisierung und wurden somit nicht für die Entwicklung von Aufgaben berücksichtigt. Daher lagen die Ausprägungen der verschiedenen Merkmale mit unterschiedlicher Häufigkeit in den Items vor und wiesen zudem signifikante Abhängigkeiten untereinander auf. Bei der Analyse von Effekten dieser Merkmale können somit unpräzise, instabile Schätzungen bzw. große Standardfehler der Schätzungen auftreten, wodurch die Teststärke eingeschränkt ist. Darüber hinaus beziehen sich Schätzungen für eine selten vertretene Merkmalsausprägung nur auf eine kleine Auswahl aller möglichen Items mit dieser Charakteristik, wodurch die Generalisierbarkeit der getroffenen Aussagen gemindert ist.

Neben der Konstruktion der Aufgabenstämme ist eine Schwäche der Itemkonstruktion im Hinblick auf die von mir untersuchte Fragestellung in der Entwicklung von Antwortalternativen bzw. deren Zuordnung zu Aufgabenstämmen zu finden. In der Construct Map für die Itementwicklung wurde versucht eine möglichst breite Auswahl gängiger Schülervorstellungen zu den untersuchten Themenbereichen zusammen zu tragen, die als Vorlage für die Formulierung von Attraktoren und Distraktoren dienen sollten. Durch diese große Auswahl und somit hohe Anzahl möglicher Antwortalternativen wurde nicht jeder Stamm mit jeder möglichen Vorstellung kombiniert, sondern jeweils nur mit einer Auswahl. Dieses Vorgehen führte u.U. zu einer Konfundierung von Schülervorstellungen und Kontextmerkmalen. Die Konfundierung ist besonders daher problematisch, da anzunehmen ist, dass auch der konkrete Inhalt und nicht allein das zugewiesene Niveau der Antwortalternativen einen bedeutsamen Einfluss auf die Aufgabenbearbeitung hat. Gerade aufgrund der Konfundierung war es anhand der Daten dieser Studie allerdings auch nicht möglich diesen Einfluss quantifizierend zu untersuchen.

Ebenso wie der bei der Entwicklung von Aufgaben wurden auch bei der Zusammenstellung von Testheften mit den Aufgaben vorrangig die postulierten Kompetenzbereiche und die untersuchten Inhaltsbereiche in den Blick genommen. Dadurch ergibt sich, dass nicht alle Ausprägungen der

Kontextmerkmale in jedem Testheft enthalten waren. Mögliche Reihenfolgeeffekte von Aufgaben mit unterschiedlichen Merkmalsausprägungen blieben ebenso völlig unberücksichtigt. Dazu wurde jedes Testheft nur von einem Teil der Stichprobe bearbeitet, so dass einige Personen nur Aufgaben mit bestimmten Charakteristika bearbeitet haben. Somit bestand eine teilweise Konfundierung von Merkmalsausprägungen und Substichproben.

Wie im vorhergehenden Abschnitt erläutert, könnte sich ein möglicher Einfluss der untersuchten Merkmale differentiell im Hinblick auf bestimmte Probandengruppen zeigen. Die von mir untersuchte Stichprobe war allerdings eher heterogen: Die Schüler stammten aus verschiedenen Klassenstufen und die Klassen hatten bereits unterschiedlich viel Unterricht in den untersuchten Themenbereichen. Mögliche Effekte könnten durch diese zusätzlichen Einflussfaktoren verschleiert worden sein.

Allgemein ist noch anzumerken, dass die empirisch ermittelten Itemschwierigkeiten sowohl auf die jeweilige Antwortalternative (beim Antwortformat *multiple select*) bzw. die jeweilige Kombination von Antwortalternativen (bei den Antwortformaten *forced choice* bzw. *multiple choice*) und den zuvor dargestellten Aufgabenstamm zurück geführt werden können. Eine konkrete Trennung der Varianzanteile von Item und Stamm wurde in den gerechneten Modellen nicht vorgenommen. Die Abhängigkeit der Items hätte mit Q3-Statistiken (Yen, 1984) untersucht werden können und eine entsprechende Varianzkomponente je Aufgabenstamm hätte modelliert werden können. Die aus solchen Modellen geschätzten Regressionskoeffizienten hätten sich dann allerdings nicht mehr auf die gleiche Grundvarianz bezogen und die Interpretation der Parameter im Sinne der untersuchten Fragestellung wäre erschwert gewesen, weshalb ich darauf verzichtet habe.

#### **4.5.3 Konsequenzen für die Aufgabenentwicklung aus Studie 1**

Trotz der erläuterten Einschränkungen, die sich aus der Anlage der Querschnittstudie ergeben haben, beeinflussten einige der untersuchten Aufgabenmerkmale signifikant die empirischen Schwierigkeiten der Items. Allerdings waren keine starken Hinweise auf die Kontextualisierung von Schülervorstellungen zu finden. Dennoch geben die Befunde dieser Studie einige Hinweise, die bei zukünftigen Aufgabenentwicklungen berücksichtigt werden sollten.

Der Verteilung der Ausprägungen der Kontextmerkmale wurde während der Itementwicklung keine Aufmerksamkeit geschenkt, weshalb die Vorkommenshäufigkeit sich teilweise extrem unterschied. Dies könnte allerdings dazu führen, dass das Instrument als Ganzes das Verständnis von beispielsweise Verdunstung in einer sehr speziellen Auswahl von Situationen erfasst und nicht in einer repräsentativen Auswahl von Situationen in denen Verdunstungsphänomene zu beobachten sind. Um eine möglichst hohe inhaltliche Validität des Instrumentes zu sichern, sollte eine genaue Analyse der Kontexte, in denen das Verständnis erfasst werden soll, als zusätzliche Grundlage der Itementwicklung berücksichtigt werden.



Vor allem für die Untersuchung der hier vorliegenden Frage nach dem systematischen Einfluss von Aufgabenmerkmalen, sind aus den Ergebnissen hilfreiche Schlüsse zu ziehen: Je nach Intention der Aufgabenentwicklung sollte eine Kontrolle bzw. systematische Variation der Aufgabenmerkmale angestrebt werden. Um möglichst frei von Störfaktoren und ohne Konfundierung mit weiteren Merkmalen die Auswirkung bestimmter Aufgabenattribute zu untersuchen, bietet dies eine gute Möglichkeit dieser Frage nachzugehen. Allerdings ist auch in dem Fall zu bedenken, dass die inhaltliche Validität eines solchen Erhebungsinstrumentes u.U. eingeschränkt sein kann. Weicht die Auswahl von Aufgabenkontexten stark von der natürlich Auftretensverteilung des untersuchten Phänomens oder den möglichen lebensweltlichen Erfahrungen der untersuchten Zielgruppe ab, könnte die Gültigkeit der Ergebnisse gemindert sein. Falls die Dimension der Kontextualität von Phänomen bei der vorhergehenden Operationalisierung des untersuchten Konstrukts berücksichtigt wurde, besteht dieses Problem allerdings nicht bzw. durch die optimale Auswahl von Aufgabensituationen könnten in diesem Fall die inhaltliche Validität sogar gesteigert werden.

Die deutlichsten Befunde zeigten sich in dieser Studie bei den Aufgaben zum Bereich Dichte. Ein großer Teil der Varianz konnte durch ein Aufgabenmerkmal, die Kongruenz, aufgeklärt werden. Die Erklärungsmacht der Kongruenz ist so hoch, dass bei weiteren Aufgabenentwicklungen zu diesem Bereich dieses Merkmal berücksichtigt werden sollten. Eine gezielte Variation der Aufgabenschwierigkeit durch regelgeleitete Aufgabenentwicklung mittels dieses Merkmals ist ebenfalls denkbar.

#### ***4.5.4 Fazit aus Studie 1 und Ausblick***

Im Hinblick auf die zentrale Fragestellung der Arbeit nach dem systematischen Einfluss von Aufgabenmerkmalen, besonders Merkmalen des Kontexts, zeigten die Ergebnisse dieser Studie ein tendenzielles Bild. In einigen Bereichen des Wissens, v.a. im Themenbereich Verdunstung/Kondensation und bei den Komponenten Dichte und Verdrängung, waren Hinweise auf eine Kontextualisierung von Schülervorstellungen zu finden.

Obwohl zwar einige der untersuchten Merkmale die Itemschwierigkeit signifikant beeinflussten, konnten bei anderen Merkmalen keine entsprechenden Effekte nachgewiesen werden. Da die untersuchten Kontextmerkmale allerdings nicht systematisch in die Aufgabenentwicklung mit einbezogen wurden, waren einige methodische Beschränkungen der Querschnittstudie festzuhalten. Dieser Umstand legt den Schluss nahe, dass die nicht-signifikanten Ergebnisse nicht nur auf ein Nicht-Wirken der untersuchten Faktoren zurückzuführen sind, sondern möglicherweise ein Artefakt der Datenlage sind.

Daher sollen in einer zweiten Studie Aufgaben aus der ersten Studie überarbeitet und dabei ausgewählte Kontextmerkmale für jeden Themenbereich experimentell variiert werden. Ziel soll es hierbei sein, systematisch Kontextmerkmale in den Aufgaben zu variieren, während weitere Merkmale konstant gehalten und somit kontrolliert werden. Dies führt dazu, dass unterschiedliche

Ausprägungen der Merkmale gleich häufig in den Aufgaben anzutreffen sind und dass keine numerischen Abhängigkeiten zwischen den Merkmalen mehr bestehen. Neben der Kontrolle zusätzlicher Kontextmerkmale sollen auch formale Merkmale konstant gehalten werden, um auch deren Einfluss zu minimieren. Darüber hinaus sollen auch die als Attraktoren bzw. Distraktoren angebotenen Schülervorstellungen gleich gehalten werden, um eine Konfundierung zwischen ihnen und den Aufgabenkontexten zu vermeiden.

Die Zusammenstellung von Aufgaben zu Testheften soll nicht im Multi-Matrix-Design erfolgen, um durch das Testheftdesign bedingte fehlende Werte und Konfundierung mit Substichproben auszuschließen. Vielmehr werden die Probanden einer Stichprobe ein vollständiges Testheft mit allen Aufgaben zu einem Bereich bearbeiten, wodurch ein experimentelles Design mit Messwiederholungen entsteht. Die Stichprobe wird in der nachfolgenden Studie einer homogenen Grundgesamtheit entnommen, so dass nicht mit differentiellen Effekten aufgrund von Jahrgangsstufe oder Alter zu rechnen ist.

## 5 Studie 2: Experimentelle Variation von Aufgabenmerkmalen

### 5.1 Einbettung der Studie 2 in das Projekt Science-P

Nach der Validierung des postulierten Kompetenzmodells in einer querschnittlichen Untersuchung wurde im Projekt Science-P eine Überführung des Stufenmodells in ein Modell der Kompetenzentwicklung angestrebt, welches in einer anschließenden Längsschnittstudie überprüft werden soll. Zur Vorbereitung dieser längsschnittlichen Erhebung wurden, da weitere formale und inhaltliche Veränderungen an den Aufgaben und Testheften vorgenommen wurden, erneut Aufgaben aus dem bisherigen Pool zur Erfassung der Kompetenz auf der Dimension *Naturwissenschaftliches Wissen* ausgewählt, überarbeitet und pilotiert. Die signifikantesten Modifikationen der Aufgaben im Projekt Science-P betrafen dabei auf inhaltlicher Ebene den Themenbereich Schwimmen/Sinken: Da in der Querschnittstudie die Aufgaben zur Komponente Auftrieb eine ungenügende psychometrische Qualität aufwiesen und es darüber hinaus aufgrund einer verringerten gesamten Aufgabenzahl in der Längsschnittstudie kritisch sein könnte drei Komponenten innerhalb eines Themenbereichs abzubilden, wurde beschlossen, auf diese Komponente in den folgenden Studien zu verzichten. Im Themenbereich Verdunstung/Kondensation wurden die Aufgaben besonders in formaler Hinsicht modifiziert: Das Antwortformat wurde für alle Aufgaben zu *multiple select* verändert und gleichzeitig die Anzahl verschiedener Antwortalternativen je Aufgabenstamm erhöht. Dadurch sollte v.a. die Erhebung einer breiteren Anzahl von *naiven Vorstellungen* möglich werden.

Es wurden drei Pilotierungsstudien, separat für Verdunstung<sup>9</sup>, Kondensation und Dichte/Verdrängung, durchgeführt. Die Anlage der Studie ermöglichte mir die Implementierung eines experimentellen Designs zur Prüfung des Effekts von Aufgabenmerkmalen.

### 5.2 Präzisierung und Einschränkung der Fragestellungen und Hypothesen für Studie 2

Die Untersuchung der Fragestellungen dieser Arbeit an dem Aufgabenpool in Studie 1 konnte nur für einige der untersuchten Merkmale einen signifikanten Einfluss auf die Aufgabenschwierigkeit nachweisen. Da v.a. die untersuchten Kontextmerkmale allerdings nicht systematisch in die Aufgabenentwicklung mit einbezogen wurden, war die Verteilung der Items hinsichtlich der untersuchten Merkmale häufig ungleich zugunsten einer Merkmalsausprägung. Dieser Umstand legt den Schluss nahe, dass die nicht-signifikanten Ergebnisse nicht nur auf ein Nicht-Wirken der

---

<sup>9</sup> Die Pilotierungstudie zur Komponente Verdunstung hat Sina Marten 2011 im Rahmen ihrer Master-Arbeit „Was ist schwerer? - Eine Untersuchung zum Einfluss von Kontextmerkmalen auf die Aufgabenschwierigkeit in einem Fragebogen zum konzeptuellen Verständnis von Verdunstung von Grundschulern“ unter meiner Betreuung am Seminar für Didaktik des Sachunterrichts durchgeführt. Die Daten wurden für die vorliegende Arbeit neu ausgewertet.

untersuchten Faktoren zurück zu führen sein können, sondern möglicherweise ein Artefakt der Datenlage sind. Daher wurden in einer zweiten Studie Aufgaben aus der ersten Studie überarbeitet und dabei ausgewählte Kontextmerkmale für jeden Themenbereich experimentell variiert. Es ergaben sich für diese Studie folgende Fragestellungen:

Da das Vorliegen der Kontextmerkmale nun als unabhängige Variable manipuliert wurde, war in dieser Studie die Frage 1<sup>10</sup> nach dem Umfang des Vorliegens der unterschiedlichen Merkmale nicht zu beantworten, sondern wurde durch das experimentelle Design vorgegeben. Die Frage 1 entfiel in Studie 2.

Sowohl Frage 2, nach dem Einfluss der Kontextmerkmale auf die Aufgabenschwierigkeit, als auch Frage 3, nach einer möglichen Interaktion zwischen Kontextmerkmalen und postuliertem Niveau einer Aufgabe, sollten in der zweiten Studie untersucht werden:

**Frage 2 (F2):**

Beeinflussen die in den Aufgaben zur Erfassung naturwissenschaftlicher Kompetenz von Grundschulern enthaltenen Kontextmerkmale das Antwortverhalten der Probanden in der Experimentalstudie und somit die Aufgabenschwierigkeit?

**Frage 2 – Hypothese 1 (F2-H1):** Die untersuchten Kontextmerkmale zeigen einen bedeutsamen Einfluss auf die Aufgabenschwierigkeit in der Experimentalstudie.

**Frage 2 – Hypothese 2 (F2-H2):** Die Richtung des Effekts der Kontextmerkmale in der Experimentalstudie entspricht der in Tabelle 1 explizierten angenommenen Wirkrichtung.

**Frage 3 (F3):**

Verändert sich der Einfluss der Kontextmerkmale auf die Aufgabenschwierigkeit in Abhängigkeit von der wissenschaftlichen Angemessenheit bzw. dem Niveau einer Aufgabe?

**Frage 3 – Hypothese 1 (F3-H1):** Die Stärke des Effekts eines Kontextmerkmals auf die Aufgabenschwierigkeit verringert sich mit zunehmendem Antwortniveau, d.h. es gibt in der Experimentalstudie signifikante Interaktionen zwischen Kontextmerkmalen und Niveaus bei Merkmalen, bei denen alle möglichen Kombinationen von Merkmal und Niveau vorliegen.

---

<sup>10</sup> Für eine bessere Übersichtlichkeit und leichtere Vergleichbarkeit der zwei Studien dieser Arbeit wird die ursprüngliche Nummerierung der Fragestellungen beibehalten, obwohl diese im Falle von Studie 2 nicht bei 1 beginnt.

Neben der systematischen Variation der interessierenden Kontextmerkmale wurden als weitere mögliche Einflussgrößen verschiedene formale Merkmale der Aufgaben in der zweiten Studie angeglichen und somit kontrolliert. Die Untersuchung der Frage 4, nach dem Einfluss formaler Aufgabenmerkmale auf die Aufgabenschwierigkeit, fiel daher in dieser Studie weg.

Neben dem Niveau der Antwortalternativen wurde für Studie 2 der konkrete Inhalt der Itemantworten systematisch variiert. Dieses Vorgehen verhinderte die Konfundierung zwischen Iteminhalt und bestimmten Ausprägungen der Kontextmerkmale. Zudem eröffnete sich aber auch die Möglichkeit den Einfluss des Inhalts auf die Itemschwierigkeiten gezielt zu untersuchen. Dies ist erstrebenswert, da anzunehmen ist, dass neben dem globalen Niveau der vorgegebenen Antworten auch der konkrete Inhalte die Aufgabenbearbeitung durch die Probanden maßgeblich beeinflusst (siehe auch Kapitel 4.5.2). Daher wurde in Studie 2 folgende Fragestellung zusätzlich untersucht:

**Zusätzliche Frage (FZ):**

Wie beeinflusst der konkrete Inhalt der vorgegebenen Antwortalternativen die Aufgabenschwierigkeiten in Studie 2?

## 5.3 Methode Studie 2

### 5.3.1 Anlage und Aufbau der Untersuchung in Studie 2

#### 5.3.1.1 Stichprobe der Studie 2

**Verdunstung** Die Aufgaben zu Verdunstung wurden von acht Klassen mit insgesamt 158 Schülern am Ende der dritten Jahrgangsstufe der Grundschule im Schuljahr 2010/2011 bearbeitet, von denen 85 männlich und 73 weiblich waren (siehe Tabelle 17). Die Probanden waren im Mittel  $M=9.80$  Jahre alt und der Großteil von ihnen (66.0%) gab an im elterlichen Haushalt ausschließlich Deutsch zu sprechen. Sechs der acht Lehrkräfte aus den an der Untersuchung teilnehmenden Klassen gaben auf eine offene Nachfrage an, das Thema Wasser(-kreislauf) oder Wetter bereits im Unterricht behandelt zu haben. Da hierbei größtenteils auch auf das Thema Aggregatzustände und ihre Übergänge Bezug genommen wurde, ist davon auszugehen, dass das Thema Verdunstung den meisten Schülern bereits durch den Unterricht bekannt war.

**Tabelle 17 Stichprobenbeschreibung Studie 2**

	N <sub>Klassen</sub>	N <sub>Schüler</sub>	Alter <sup>a</sup>	Geschlecht <sup>b</sup>	Sprache <sup>c</sup>	Bücher <sup>d</sup>
Verdunstung	8	158	9.80(0.54)	85/73	66.0/34.0	-
Kondensation	4	77	10.19(0.58)	34/43	51.9/48.1	3.18
Dichte/Verdrängung	4	73	9.82(0.55) <sup>e</sup>	37/36	75.3/24.7	3.78

---

*Anmerkungen.* <sup>a</sup> M(SD) <sup>b</sup>Anzahl Jungen/Anzahl Mädchen/Anzahl fehlende Angaben <sup>c</sup>Schülerangabe zum häuslichen Sprachgebrauch in Prozent: zu Hause wird nur deutsch gesprochen/zu Hause wird noch eine andere Sprache gesprochen (fehlende Prozente sind fehlende Angaben) (vgl. Lenhard & Schneider, 2006) <sup>d</sup>Mittelwert der Schülerangabe zum familiären Buchbesitz: fünfstufige Likert-Skala mit den Endpunkten „keine oder nur sehr wenige Bücher (0-10)“ bis „genug, um drei oder mehr Regale zu füllen (über 200)“ (vgl. Bos et al., 2005). <sup>e</sup>Die Angabe zum Alter der Schüler wurde über einen Lehrerfragebogen erhoben, der von einer Lehrkraft der Studie zu den Bereichen Dichte und Verdrängung nicht ausgefüllt zurückgegeben wurde. Die Angaben zum Alter beruhen auf den Werten von 52 Schülern.

**Kondensation** 77 Schüler aus vier Klassen in der Mitte der vierten Klassenstufe im Schuljahr 2011/2012 nahmen an der Studie zu Kondensation teil (siehe Tabelle 17). Unter ihnen waren 34 Jungen und 43 Mädchen, die durchschnittlich  $M=10.19$  Jahre alt waren. Der Anteil an Schülern, die angaben zu Hause noch eine oder mehrere andere Sprachen als deutsch zu sprechen, war in dieser Stichprobe mit 48.1% vergleichsweise hoch. Auch die Lehrer dieser Klassen wurden gebeten auf einem kurzen Fragebogen Angaben zum Unterricht in dem untersuchten Themenbereich zu machen, was allerdings nur von zwei der vier Lehrkräfte wahrgenommen wurde. Diese gaben an, bereits eine oder noch keine Doppelstunde lang das Thema Kondensation im Unterricht behandelt zu haben.

**Dichte/Verdrängung** Das Testheft mit den Aufgaben zu Dichte und Verdrängung wurden von vier Klassen zu Beginn des vierten Schuljahres der Grundschule im Jahr 2011/2012 bearbeitet (siehe Tabelle 17). Somit nahmen 73 Probanden mit einem durchschnittlichen Alter von  $M=9.82$  Jahren an der Untersuchung teil, von denen 37 männlich und 36 weiblich waren. 75.3% der Schüler gaben an, zu Hause keine andere Sprache als deutsch zu sprechen. Es liegen keine Informationen zu möglichem Unterricht in dem untersuchten Bereich aus diesen vier Klassen vor.

#### *5.3.1.2 Instrumente zur Erfassung des Naturwissenschaftlichen Wissens in Studie 2*

**Verdunstung** Für den Bereich Verdunstung wurden 24 Aufgaben im Format *multiple select* mit jeweils sechs Antwortalternativen entwickelt. Von den Antworten waren drei dem naiven, zwei dem mittleren und eine dem wissenschaftlichen Niveau zugeordnet. Die Konzepte, die in den Antwortalternativen als Erklärungen für die dargestellten Situationen in den Aufgaben gegeben waren, waren über alle Aufgaben hinweg identisch, wurden jedoch in bis zu vier unterschiedlichen, aber inhaltlich äquivalenten Formulierungen dargeboten. Die konkreten inhaltlichen Konzepte wurden ausgewählt nach der Häufigkeit ihrer Auswahl durch die Probanden in vorhergehenden Studien und dem Vorkommen in Interviews mit Probanden der gleichen Altersgruppe (vgl. Pollmeier et al., 2011). Darüber hinaus wurden außerdem psychometrische Eigenschaften und die Plausibilität im Zusammenhang mit den ausgewählten Aufgabenstämmen berücksichtigt. Die Antwort des wissenschaftlichen Niveaus beschrieb den Prozess der Verdunstung unter Berücksichtigung des Wechsels des Aggregatzustandes, der Massenerhaltung und des Ortswechsel (wissenschaftsnähe Vorstellung). Bei den zwei Antwortalternativen auf dem mittleren Niveau wurde jeweils entweder der Orts- oder der Zustandswechsel aufgegriffen, wobei die

Massenerhaltung impliziert war. Auf dem naiven Niveau wurden als Antwortalternativen eine falsche Beschreibung der Zustandsänderung (Wasser wird zu etwas immateriellem) und eine anthropomorphe Deutung der Verdunstung, bei der ein aktiver Agent das Wasser zu sich zieht, angeboten. Als aktiver Agent wurde dabei teilweise die Sonne<sup>11</sup> und teilweise die Luft gewählt. Die dritte Antwortalternative auf dem unteren Niveau sollte eine Ad-hoc-Konstruktion sein, die sich jeweils auf Spezifika der Situation bezieht, da sich Grundschüler häufig auf einzelne Aspekte von Situationen bzw. Kontexten beziehen, um eigene Antworten zu generieren. Die Konstruktion von entsprechenden plausiblen Antwortalternativen war allerdings nicht immer unproblematisch, daher wurde bei 8 von 24 Aufgabenstämmen statt der Ad-hoc-Konstruktion eine Antwort, die die Nichterhaltung der Masse thematisiert, angeboten.

**Tabelle 18 Verteilung der Aufgaben zu Verdunstung auf die Kontextmerkmale in Studie 2**

		Wasserspiegel					
		ja			nein		
		Temperatur			Temperatur		
		warm	Sonne	kalt	warm	Sonne	kalt
Ort	drinnen	b3	b5	b1	b9	b11	b7
		b15	b17	b13	b21	b23	b19
	draußen	b4	b6	b2	b10	b12	b8
		b16	b18	b14	b22	b24	b20

In den Aufgabenstämmen variierten die drei Kontextfaktoren Sichtbarkeit des Wasserspiegels (zweifach gestuft), Ort (zweifach gestuft) und Temperatur (dreifach gestuft) systematisch. Jede von zwölf Zellen in dem so entstehenden experimentellen Design war mit zwei Aufgaben mit je sechs Items besetzt (siehe Tabelle 18). Dabei wurden teilweise parallele Aufgabenstämme realisiert, die sich nur hinsichtlich eines der Kontextmerkmale unterschieden. Beispielsweise wurde das Verdunsten von Wasser aus einem Glas einmal innerhalb (drinnen) und einmal außerhalb (draußen) geschlossener Räume beschrieben.

Für die Zusammenstellung der Aufgaben zu Testheften wurde die Reihenfolge der Antwortalternativen zu jedem Aufgabenstamm variiert und die Darbietungsabfolge der Aufgaben war in jeder teilnehmenden Klasse unterschiedlich. Auf der ersten Seite jedes Testheftes wurden zudem das Geschlecht und Geburtsdatum der Schüler und der häusliche Sprachgebrauch erfragt.<sup>12</sup>

**Kondensation** Die Aufgaben zu Kondensation wurden analog zu den Verdunstungs-Aufgaben überarbeitet (MS-Format mit sechs Antwortalternativen). Die Antwortalternativen auf dem wissenschaftlichen und dem mittleren Niveau entsprachen den Antwortalternativen, die auch zum

<sup>11</sup> Sonne war in dieser Studie, wie auch in Studie 1, eine Merkmalsausprägung des Kontextmerkmals Temperatur und sollte daher bewusst nur in einer bestimmten Auswahl von Aufgaben explizit erwähnt werden.

<sup>12</sup> Alle Testhefte der Studie 2 sind in der Version für Testleiter in den elektronischen Anhängen zu finden. Dabei wurde jeweils nur eine mögliche Reihenfolge der Aufgaben dargestellt.

Für den Bereich Kondensation wurde nur ein Testheft zusammengestellt, das alle 19 Aufgaben in jeder Klasse in der gleichen Reihenfolge darbot. Die Reihenfolge der Antwortalternativen wurde für jeden Aufgabenstamm variiert.

			Quelle			
			vorhanden		nicht vorhanden	
			Temperatur		Temperatur	
			warm	kalt	warm	kalt
Kk	Fläche	drinnen	k05	k03	k01	k06
			k10	k13	k19	k16
			k15			
	draußen	k02	k09	k08	k11 k18	
Dampf	Ort	drinnen	k17	---	---	---
		draußen	k07	k12	---	k04 k14

Aufgrund der aufgetretenen Schwierigkeiten bei der Konstruktion von Aufgabenstämmen wurden die Probanden zudem nach jeder Aufgabe gebeten die Alltagsnähe der Situation auf einer vierstufigen Likert-Skala zu beurteilen. Auf dem Deckblatt wurde darüber hinaus das Geschlecht, der häusliche Sprachgebrauch sowie die Anzahl der Bücher im elterlichen Haushalt erfragt.

Die Aufgaben zu Dichte wurden in ihrem Format (*multiple choice* mit drei Antwortalternativen) nicht verändert. In diesen Aufgaben wurde das Merkmal der Kongruenz systematisch variiert.

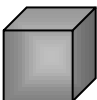


Insgesamt wurden so zwölf Aufgaben zum Dichtekonzept entwickelt, von denen drei als kongruent und drei als inkongruent einzustufen sind (siehe Tabelle 20). Darüber hinaus gab es je drei Aufgaben, in denen entweder das Gewicht der zu vergleichenden Objekte oder das Material bzw. die spezifische Dichte der Objekte konstant gehalten wurde. Eine Beispielaufgabe mit inkongruenten Objekten (d.h. der Klotz mit dem geringeren Gewicht ist aus dem schwereren Material gemacht) ist in Abbildung 4 zu sehen.

**Abbildung 4 Beispielaufgabe zum Bereich Dichte, Aufgabentyp inkongruent**


**Klötze vergleichen**  
*Instruktion des Testleiters:*  
 Ihr seht hier zwei Klötze. Klotz 1 wiegt 10 Kilogramm. Klotz 2 wiegt 12 Kilogramm.

Klotz 1:



10 Kilogramm

Klotz 2:



12 Kilogramm

Welcher Klotz ist aus dem schwereren Material gemacht?

Kreuze die beste Antwort an!

Klotz 1	<input type="checkbox"/>
Klotz 2	<input type="checkbox"/>
Beide sind aus dem gleichen Material.	<input type="checkbox"/>

Die kürzeste Skala war mit nur sechs Aufgaben im Instrument der Querschnittstudie die Skala zum Bereich Verdrängung. Auch die Untersuchung des Einflusses der Kontextfaktoren war durch die geringe Anzahl der Faktoren eingeschränkt. Der Unterschied zwischen dem untersuchten naiven Konzept in diesem Bereich (Verdrängung ist vom Gewicht abhängig) und dem wissenschaftsnahen Konzept (Verdrängung ist vom Volumen abhängig) ähnelt stark den im Bereich Dichte untersuchten Konzepten (Dichte und Gewicht werden nicht differenziert vs. Dichte ergibt sich aus Masse und Volumen eines Objektes). Daher wurden für Verdrängung sechs völlig neue Aufgaben entwickelt, die in ihrem äußeren Design den Dichte-Aufgaben nachempfunden waren. Auch in diesen Aufgaben wurde die Kongruenz variiert (siehe Tabelle 20). Bei kongruenten Aufgaben war der schwerere Gegenstand auch der mit dem größeren Volumen, der also mehr Wasser verdrängt. Bei inkongruenten Aufgaben hatte der größere Gegenstand das geringere Gewicht. Zudem gab es Aufgaben in denen sich die zu vergleichenden Gegenstände nur in ihrem Volumen, aber nicht in ihrem Gewicht unterschieden.

Das Testheft zum Bereich Dichte/Verdrängung umfasste insgesamt 22 Aufgaben. Darin wurden die beschriebenen zwölf Aufgaben zu Dichte und sechs Aufgaben zu Verdrängung gemeinsam mit vier Verdrängungsaufgaben aus der Querschnittstudie dargeboten.<sup>13</sup> Die Erhebung demographischer Daten erfolgte wie in der Studie zum Bereich Kondensation.

**Tabelle 20 Verteilung der Aufgaben zu Dichte und Verdrängung auf das Kontextmerkmal in Studie 2**

		Dichte	Verdrängung
Kongruenz	inkongruent	d7	v3
		d8	v4
		d9	
	gleiches Material	d4	
		d5	---
		d6	
	gleiches Gewicht	d1	v5
		d2	v6
		d3	
	kongruent	d10	v1
		d11	v2
		d12	

#### 5.3.1.3 Weitere Erhebungsinstrumente in Studie 2

Zusätzlich zum Instrument für die Schüler wurde in den Erhebungen zu Kondensation und Dichte/Verdrängung ein kurzer Fragebogen an die Lehrkräfte ausgegeben, in dem diese Angaben zum Geburtsdatum der Kinder und zum erteilten Unterricht in den untersuchten Themenbereichen machen konnten.

#### 5.3.1.4 Durchführung der Studie 2

Die Durchführung der Aufgaben als Gruppentest in den Schulklassen entsprach dem in Studie 1 beschriebenen Vorgehen mit Testleiter, Testassistenten und Projektion der Aufgaben in einer Unterrichtsdoppelstunde von 90 Minuten. Allerdings wurde in Studie 2 auf eine Demonstration der in den Aufgaben beschriebenen Situationen mit realen Gegenständen verzichtet, um so einen weiteren möglichen Einflussfaktor auf die Schwierigkeiten der Aufgaben konstant zu halten. Entsprechend der Ergebnisse aus Studie 1 zum diesem Faktor werden dadurch keine großen Veränderungen in den Schwierigkeiten der Aufgaben erwartet.

Eine zusätzliche Instruktion in Studie 2 bezog sich auf die wiederholte Darbietung inhaltlich äquivalenter Antwortalternativen: Die Schüler wurden zu Beginn der Testung darauf hingewiesen, dass sie bei jeder Aufgabe neu überlegen sollten und dann eine Antwort auswählen sollten. Kam es in einigen vereinzelt Fällen zu Nachfragen der Schüler bezüglich dieses Aspektes, wurden die

<sup>13</sup> Die Durchführung dieser Studie diente im Rahmen des Projekts Science-P u.a. zur Überprüfung einer möglichen Eindimensionalität des Inhaltsbereichs Dichte/Verdrängung ohne Aufgaben zum Auftrieb. Daher wurden auch die Aufgaben der Querschnittstudie in diesem Testheft eingesetzt.

Schüler erneut darauf hingewiesen, dass sie über die Antworten im Zusammenhang mit der neuen Aufgabe erneut nachdenken sollten.

### **5.3.2 Datenaufbereitung und –auswertung in Studie 2**

#### *5.3.2.1 Kodierung der Schülerantworten auf die Aufgaben in Studie 2*

Wie auch in Studie 1 wurden in dieser Studie alle Items, die entweder ein MC- oder MS-Format aufwiesen, dichotom kodiert. Dabei wurde bei MC-Aufgaben die Auswahl der höchsten vorgegebenen Niveaus bzw. der richtigen Antwort bepunktet. Bei MS-Aufgaben wurde bei Antwortalternativen auf dem naiven Niveau die Ablehnung, bei Antworten auf dem mittleren und wissenschaftlichen Niveau die Zustimmung bepunktet.

#### *5.3.2.2 Kodierung der Aufgaben hinsichtlich der untersuchten Merkmale in Studie 2*

Wie schon in Studie 1 wurde jedes Item nur einer Ausprägung eines Aufgabenmerkmals zugewiesen. Die Zuordnung zu den Ausprägungen der Kontextmerkmale war bereits vorab durch die Zuweisung im experimentellen Design der Studie geleistet.

Die Kodierung der Items hinsichtlich der Kompetenzniveaus wurde wieder anhand des höchsten erreichbaren Niveaus vorgenommen. Dem höchsten Niveau der wissenschaftlichen Vorstellungen wurde z.B. eine MC-Aufgabe, in der alle drei Niveaus als Alternativen angeboten wurden, zugeordnet. Die Antwortalternativen von MS-Aufgaben wurden dem jeweiligen Niveau zugeordnet, dem das inhaltliche Konzept, welches sie als Erklärung anboten, in der Construct Map zugeordnet war. Die Zuordnung entsprach damit der in Studie 1 (siehe Kapitel 4.3.2.2).

Für die Bereiche Verdunstung und Kondensation wurden darüber hinaus die inhaltlichen Konzepte, die als Antworten angeboten wurden, selbst kodiert, was durch die Wiederholung der Alternativen über Aufgaben hinweg sinnvoll möglich wurde. Dazu wurden die Konzepte zur Identifizierung von 1 bis 6 nummeriert. Diese Variable kann als Faktor in den Analysen verwendet werden.

#### *5.3.2.3 Dimensionalitätsprüfung, Itemselektion und zentrale psychometrische Kennwerte der Skalen und Items in Studie 2*

Auch in Studie zwei wurden die Daten vor den Analysen zur Überprüfung der Hypothesen einer Analyse ihrer psychometrischen Qualität mit dem Programm ACER Conquest (Wu, Adam & Wilson, 2005) unterzogen. Auf eine Analyse der Dimensionalität der Daten wurde in dieser Studie verzichtet, da keine weiteren inhaltlichen Untergliederungen der Aufgaben durch die Construct Map angenommen wurden.

Es wurde für jeden Bereich ein eindimensionales, einparametriges logistisches Modell an die Daten angepasst und die psychometrischen Eigenschaften der Items wurden anhand dieser Skalierungen überprüft. Die Auswahl von Items erfolgte wie auch in Studie 1 anhand der gewichteten Abweichungsquadrate (*weighted mean square*, MNSQ; Ausschluss wenn MNSQ > 1.30), der

Trennschärfe (Ausschluss wenn  $r_{it} < .10$ ) und dem Vorliegen von Boden- und Deckeneffekten (Ausschluss wenn Anzahl korrekter Antworten  $< .10 | > .90$ ). Eine detaillierte Darstellung der psychometrischen Kennwerte aller Items ist in Anhang A07 zu finden.

**Verdunstung** Nach der ersten eindimensionalen Skalierung der 144 Items zu diesem Konzept zeigten 25 Items ungenügende psychometrische Qualität und wurden von den weiteren Analysen ausgeschlossen. Dabei war der Grund der Selektion jedes Mal eine nicht ausreichende Trennschärfe. Auffällig an den ausgeschlossenen Items war, dass ausschließlich zwei der sechs möglichen Antwortalternativen ausgeschlossen werden mussten: 19-mal wurde die Antwortalternative anthropomorphe Deutung ausgeschlossen und 6-mal die Antwort, bei der es um einen ad-hoc Konstruktion bzw. Nicht-Erhaltung der Masse ging. Dabei war noch einmal besonders auffallend, dass die anthropomorphe Deutung besonders selten in Kontexten ausgeschlossen werden musste, in denen die Sonne als Wärmequelle genannt wird. Die Skalierung der verbleibenden 119 Items wies eine Reliabilität von 0.94 sowie eine Varianz von 0.78 auf.

**Kondensation** Die Skalierung und Auswahl der Kondensationsitems stellte sich als problematisch dar. Es mussten insgesamt 59 Items, größtenteils wegen mangelnder Trennschärfe, in sieben Fällen wegen mangelnder Passung, aus dem Datensatz ausgeschlossen werden, der nur noch 55 der ursprünglich 114 Items enthält. Von den selektierten Items waren 5 dem mittleren und 53 dem naiven Niveau zuzuordnen. Nur ein ausgeschlossenes Item gehörte dem wissenschaftlichen Niveau an. Es ist also bei der Skala zu Kondensation nicht gelungen Antwortalternativen zu formulieren, die eine sinnvolle Erhebung der naiven Vorstellungen in diesem Bereich ermöglichten. Davon wird die spätere Interpretation der Ergebnisse betroffen sein, da diese Skala nun eher als ein Abfragen von fortgeschritteneren Vorstellungen beschrieben werden kann. Eine Reliabilität von 0.92 und eine Varianz von 1.31 erzielte die Skalierung der verbleibenden 55 Items.

**Dichte/Verdrängung** Trotz der Darbietung der Aufgaben zu diesen beiden Bereichen in einem Testheft werden die Items zu Dichte und Verdrängung für Studie 2 separat ausgewertet. Diese beiden Skalen sind deutlich kürzer als die Skalen zu Verdunstung und Kondensation, was einerseits auf das MC-Antwortformat zurück zu führen ist, bei dem nach jedem Aufgabenstamm nur eine Antwortreaktion von den Probanden verlangt wird. Andererseits wurden diese Bereiche im Projekt Science-P weniger stark gewichtet als andere Bereiche, was ebenfalls zu einer kürzeren Skala führte.

Von den zwölf Dichte-Aufgaben wurden zwei wegen ungenügender Passung von den weiteren Analysen ausgeschlossen. Dabei ging es in einer Aufgaben um den Vergleich der Dichte eines inkongruenten Gegenstandspaares und in einer Aufgabe hatten die Gegenstände das gleiche Gewicht. Bei der Analyse der verbleibenden 10 Items zeigten sich eine Varianz von 3.98 und eine Reliabilität von 0.73.

Für den Bereich Verdrängung war der Ausschluss von Aufgaben aufgrund der Kürze der Skala, die eine sehr niedrige Reliabilität zur Folge hatte, problematisch. Eine der sechs Aufgaben zeigte eine

ungenügende Passung. Die Skalierung der verbleibenden 5 Items hatte zwar eine Varianz von 5.02. Die Reliabilität von 0.17 lag allerdings in einem ungenügenden Bereich. Die Skalierung dieser Aufgaben zusammen mit den Verdrängungsaufgaben aus Studie 1 (WLE-Reliabilität=.78, Varianz=9.97), sowie eine Überprüfung der Reliabilität mit Methoden der klassischen Testtheorie zeigten zufriedenstellende Kennwerte (Cronbach's  $\alpha$ =.72). Auch funktionierten die für Studie 2 konstruierten Aufgaben im Verbund mit den bewährten Aufgaben aus Studie 1 gut und zeigten hinreichende Kennwerte. Aufgrund dieser Befunde ging ich davon aus, dass die Items trotz der geringen Reliabilität zur Untersuchung der vorliegenden Fragestellungen geeignet sind und die ungenügenden Kennwerte v.a. auf die geringe Anzahl an Aufgaben zurück zu führen waren. Die Skala könnte in dieser Form allerdings nicht zur Diagnostik genutzt werden.

#### *5.3.2.4 Vorgehen und Verfahren zur Hypothesenprüfung in Studie 2: Generalisierte lineare gemischte Modelle mit gekreuzten zufälligen Effekten*

Wie auch in Studie 1 sollte in Studie 2 der Einfluss der Aufgabenmerkmale mit Hilfe einer Erweiterung des linear-logistischen Testmodells untersucht werden (siehe Kapitel 4.3.2.4). Die allgemeine Formulierung der Modelle, mit einem zufälligen Effekt für Items, Personen und Klassen wurde in dieser Studie beibehalten. Die Interpretation der resultierenden Parameter (Y-Achsenabschnitt, Regressionskoeffizienten) blieb also unverändert. Die Berechnung der Modelle wurde erneut mit dem Softwarepaket lme4 in R durchgeführt (Doran et al., 2007; Baayen et al., 2008; De Boeck et al., 2011; De Boeck & Wilson, 2004). Es wurden die gleichen Kriterien (BIC: Schwartz, 1978; Chi-Quadrat-Test: Anderson, 1973) für Modellvergleiche herangezogen.

Wieder wurde zur Untersuchung der Fragestellungen zunächst der Einfluss jedes Aufgabenmerkmals konsekutiv geprüft. Aufgrund der strikt systematischen Aufgabenentwicklung im Bereich Verdunstung, durch die Abhängigkeiten der Prädiktoren untereinander ausgeschlossen sind, wurden für diesen Bereich außerdem mögliche Interaktionen der Kontextmerkmale untereinander untersucht. Dazu wurden die drei Merkmale (Sichtbarkeit des Wasserspiegels, Temperatur, Ort) gemeinsam in ein Modell eingeführt. Dabei wurden zunächst nur die Haupteffekte, im weiteren Verlauf die Interaktionen erster Ordnung sowie die Interaktion zweiter Ordnung untersucht. Mit Hilfe von hierarchischen Modellvergleichen erfolgte die Auswahl eines Optimalmodells unter diesen Modellen für Verdunstung.

Anschließend wurden auch in dieser Studie wo möglich Interaktionen der Kontextmerkmale mit den jeweiligen Niveaus der Items wie auch in der Querschnittstudie untersucht.

In den folgenden Abschnitten werden zunächst deskriptive Ergebnisse vorgestellt (siehe Kapitel 5.4.1), bevor anschließend die Befunde der inferenzstatistischen Analysen beschrieben werden (siehe Kapitel 5.4.2). Auf eine zusätzliche Aufgabenanalyse, wie sie in Kapitel 4.4.1 für Studie 1 beschrieben wurde, wird an dieser Stelle verzichtet, da durch das design-geleitete Vorgehen bei der Aufgabenentwicklung die Verteilung der Merkmale bereits a-priori festlag.

## 5.4 Ergebnisse von Studie 2

### 5.4.1 Deskriptive Ergebnisse von Studie 2

#### 5.4.1.1 Verdunstung

Da der Einfluss formaler Merkmale in dieser Studie nicht untersucht wurde, sondern diese schon während der Itemkonstruktion konstant gehalten wurden, beginnt die Ergebnisdarstellung mit den Befunden zum Einfluss der Kompetenzniveaus. Von diesen wurde das naive Niveau am häufigsten richtig gelöst ( $M=.49$ ,  $SD=.15$ ), d.h. abgelehnt, während sowohl dem mittleren ( $M=.42$ ,  $SD=.14$ ) wie auch dem wissenschaftlichen Niveau ( $M=.40$ ,  $SD=.08$ ) seltener zugestimmt wurde (vgl. Tabelle 21).

Deskriptiv ließen sich bei den Mittelwerten der Items zu den unterschiedlichen Ausprägungen der Kontextmerkmale nur geringe bzw. keine Unterschiede feststellen: Aufgaben, in denen kein Wasserspiegel sichtbar war ( $M=.43$ ,  $SD=.14$ ), wurden etwas seltener gelöst als Aufgaben, in denen er sichtbar war ( $M=.46$ ,  $SD=.14$ ). Es zeigte sich kein Unterschied in den Lösungshäufigkeiten für Aufgaben, die drinnen ( $M=.44$ ,  $SD=.14$ ) oder draußen ( $M=.44$ ,  $SD=.14$ ) situiert waren. Aufgaben, in denen keine explizite Wärme- bzw. Energiequelle ( $M=.42$ ,  $SD=.13$ ) genannt wurde, waren etwas schwieriger zu lösen als Aufgaben mit Wärmequelle ( $M=.45$ ,  $SD=.15$ ) bzw. in denen die Sonne als Energiequelle ( $M=.46$ ,  $SD=.14$ ) erwähnt wurde.

**Tabelle 21 Deskriptive Statistiken der Items zu Verdunstung in Studie 2**

	N	M	SD
Alle Items	119	.44	.14
Niveaus			
NV	47	.49	.15
ZV	48	.42	.14
WV	24	.40	.08
sichtbarer Wasserspiegel			
ja	57	.46	.14
nein	62	.43	.14
Ort			
drinnen	61	.44	.14
draußen	58	.44	.14
Temperatur			
warm	37	.45	.15
Sonne	39	.46	.14
kalt	43	.42	.13
Antwortalternative			
falsche Zustandsänderung	24	.56	.08
anthropomorphe Deutung	5	.27	.04
ad-hoc Konstruktion/keine Massenerhaltung	18	.44	.16
Ortswechsel	24	.54	.10
Zustandsänderung	24	.31	.05
wissenschaftsnähe Vorstellung	24	.40	.08

Neben den Merkmalen der Aufgabenstämme wurden die unterschiedlichen Inhalte der Antwortalternativen bzw. Items betrachtet. Dabei wurden Items, in denen die Verdunstungssituation anthropomorph gedeutet wurde (anthropomorphe Deutung;  $M=.27$ ,  $SD=.04$ ), und Items, in denen der Wechsel des Aggregatzustandes beschrieben wurde (Zustandsänderung;  $M=.31$ ,  $SD=.04$ ), am seltensten gelöst. Am häufigsten wurden hingegen Items gelöst, die die Zustandsänderung falsch ( $M=.56$ ,  $SD=.08$ ) oder den Ortswechsel korrekt ( $M=.54$ ,  $SD=.10$ ) beschrieben. Dazwischen lagen die Lösungshäufigkeiten der Items, die eine ad-hoc-Konstruktion bzw. die Nichterhaltung der Masse ( $M=.44$ ,  $SD=.16$ ) oder eine wissenschaftsnahe Vorstellung ( $M=.40$ ,  $SD=.08$ ) als Erklärung anboten.

#### 5.4.1.2 Kondensation

Bei Kondensation waren die Items des naiven Niveaus am schwierigsten zu lösen ( $M=.17$ ,  $SD=.08$ ), während die Items des mittleren ( $M=.44$ ,  $SD=.15$ ) und des wissenschaftlichen Niveaus ( $M=.48$ ,  $SD=.10$ ) deutlich häufiger richtig bearbeitet wurden (vgl. Tabelle 22).

**Tabelle 22 Deskriptive Statistiken der Items zu Kondensation in Studie 2**

	N	M	SD
Alle Items	55	.44	.15
Niveaus			
NV	4	.17	.08
ZV	33	.44	.15
WV	18	.48	.10
Ort			
drinnen	27	.44	.15
draußen	27	.44	.15
Temperatur			
kalt	24	.46	.13
warm	31	.42	.17
Kondensationskern			
Fläche	42	.44	.15
Dampf	13	.44	.16
Quelle			
vorhanden	32	.40	.15
nicht vorhanden	23	.49	.13
Antwortalternative			
falsche Ursache	1	.13	-
falsche Zustandsänderung	3	.19	.09
Ortswechsel	18	.42	.10
Zustandsänderung	15	.48	.20
wissenschaftsnahe Vorstellung	18	.48	.10

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen;

Auch bei diesen Items unterschieden sich die Mittelwerte für drinnen ( $M=.44$ ,  $SD=.15$ ) oder draußen ( $M=.44$ ,  $SD=.15$ ) situierte Aufgaben nicht. Ebenso gab es keinen Unterschied zwischen Aufgaben, in denen das Wasser auf Flächen ( $M=.44$ ,  $SD=.15$ ) oder als Dampf ( $M=.44$ ,  $SD=.16$ )

kondensierte. Die Aufgaben, in denen eine erhöhte Umgebungstemperatur ( $M=.42$ ,  $SD=.17$ ) vorlag, wurden seltener gelöst als Aufgaben mit kalter bzw. nicht erhöhter Umgebungstemperatur ( $M=.46$ ,  $SD=.13$ ). Darüber hinaus wurden die Items häufiger gelöst, wenn keine potentielle Wasserquelle in der Beschreibung zu finden war ( $M=.49$ ,  $SD=.13$ ), als die Items, in denen eine solche Quelle beschrieben wurde ( $M=.40$ ,  $SD=.15$ ).

Bei den inhaltlichen Konzepten der Antwortalternativen zeigte sich, dass es für die Probanden am schwierigsten war, die Items, die eine falsche Zustandsänderung ( $M=.19$ ,  $SD=.09$ ) oder eine falsche Ursache ( $M=.13$ ) als Erklärung anboten, zu bearbeiten. Dagegen waren sowohl die Items mit der wissenschaftsnahen Vorstellung ( $M=.48$ ,  $SD=.10$ ) als auch mit einer richtigen Beschreibung der Zustandsänderung ( $M=.48$ ,  $SD=.20$ ) leichter zu bearbeiten. Die Items, die eine richtige Beschreibung des Ortswechsels enthielten ( $M=.42$ ,  $SD=.10$ ), lagen im mittleren Bereich.

#### 5.4.1.3 Dichte und Verdrängung

Für die Items zu diesen beiden Bereichen wurde ausschließlich der Einfluss des Merkmals Kongruenz untersucht (vgl. Tabelle 23).

**Tabelle 23 Deskriptive Statistiken der Items zu Dichte und Verdrängung in Studie 2**

	Dichte		Verdrängung	
	N	M/SD	N	M/SD
Alle Items	10	.59/.21	5	.77/.12
Kongruenz				
inkongruent	2	.47/.06	2	.70/.02
gleiches Material	3	.41/.15	-	-
gleiches Gewicht	2	.59/.00	2	.73/.07
kongruent	3	.86/.05	1	.97/-

*Anmerkungen.* NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen;

Bei Dichte konnten die kongruenten Items am häufigsten korrekt bearbeitet werden ( $M=.86$ ,  $SD=.05$ ). Am schwierigsten fiel die Bearbeitung den Probanden bei Aufgaben, in denen die Gegenstände die gleiche Dichte hatten bzw. aus dem gleichen Material gemacht waren ( $M=.41$ ,  $SD=.15$ ) und inkongruenten Aufgaben ( $M=.47$ ,  $SD=.06$ ). Aufgaben, in denen das Gewicht der Gegenstände konstant gehalten wurde ( $M=.59$ ,  $SD=.00$ ), lagen dazwischen.

Die Aufgaben zu Verdrängung wurden im Mittel insgesamt deutlich häufiger gelöst als die Aufgaben zu Dichte. Auch hier war der Mittelwert der kongruenten Aufgaben ( $M=.97$ ) am höchsten. Sowohl inkongruente Aufgaben ( $M=.70$ ,  $SD=.02$ ) als auch Aufgaben, in denen die Gegenstände das gleiche Gewicht hatten ( $M=.73$ ,  $SD=.07$ ), wurden seltener richtig gelöst.



## 5.4.2 Ergebnisse der inferenzstatistischen Analysen von Studie 2

### 5.4.2.1 Verdunstung

Die Prüfung des Effekts der Aufgabenmerkmale auf die Aufgabenschwierigkeit wurde im Bereich Verdunstung für die Kontextmerkmale Wasserspiegel, Ort und Temperatur, sowie darüber hinaus für die Kompetenzniveaus und die verschiedenen Antwortalternativen vorgenommen (siehe Tabelle 24).

**Tabelle 24 Ergebnisse der Prüfung des Einflusses der Aufgabenmerkmale für Verdunstung**

	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	R <sup>2</sup>
Nullmodell	-0.30*	0.14	0.65	0.80	0.31	22724	
Niveaus			0.63	0.80	0.31	22736	.07
(NV)	-0.09	0.16					
ZV	-0.31*	0.13					
WV	-0.40*	0.16					
Wasserspiegel			0.65	0.80	0.31	22693	.01
(ja)	-0.23	0.16					
nein	-0.13	0.12					
Ort			0.65	0.80	0.31	22694	.00
(drinne)	-0.30	0.15					
draußen	0.01	0.12					
Temperatur			0.65	0.80	0.31	22695	.01
(warm)	-0.25	0.17					
Sonne	0.02	0.15					
kalt	-0.14	0.15					
Modell Kontext <sup>a</sup>			0.64	0.80	0.31	22698	.02
(Intercept)	-0.19	0.19					
Wassersp.=nein	-0.13	0.12					
Ort=draußen	0.02	0.12					
Temperatur=So	0.02	0.15					
Temperatur=kalt	-0.15	0.15					
Antwortalternative			0.42	0.80	0.31	22611	.58
(falsche Zust.)	0.29 <sup>+</sup>	0.16					
Anthropom. D.	-1.46***	0.23					
ad-hoc/Nichterh.	-0.59***	0.14					
Ortswechsel	-0.12	0.13					
Zustandsänderung	-1.26***	0.13					
wissenschaftsnah	-0.78***	0.13					

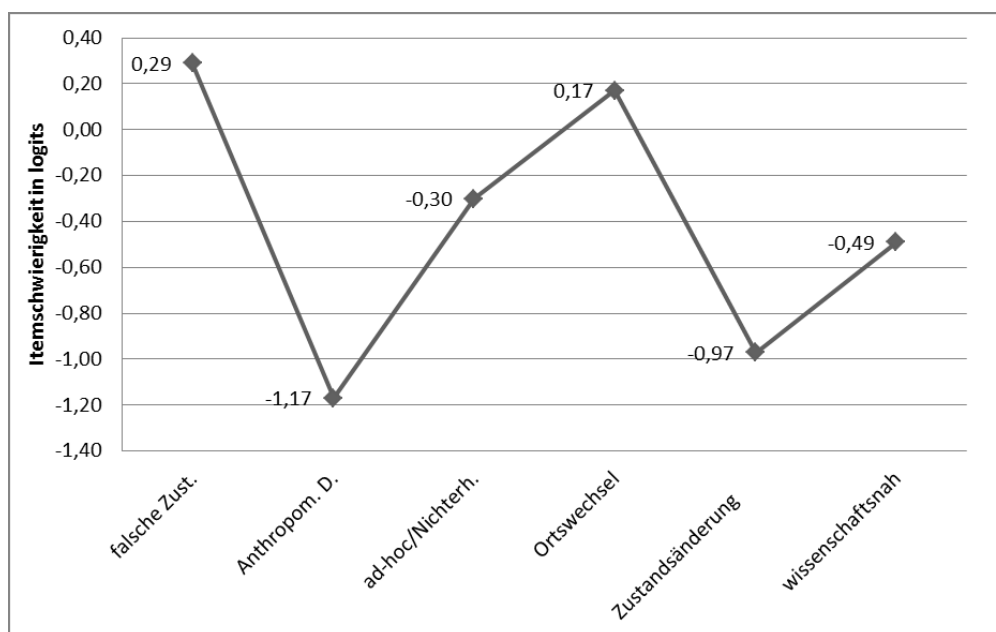
*Anmerkungen.* NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; So, Sonne; Wassersp., sichtbarer Wasserspiegel; falsche Zust., falsche Zustandsänderung; ad-hoc, Ad-hoc-Konstruktion; Nichterh., keine Massenerhaltung; wissenschaftsnah, wissenschaftsnahe Vorstellung; <sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001; <sup>a</sup>Dies ist das Modell, welches unter Berücksichtigung aller untersuchten Kontextmerkmale, in den hierarchischen Modellvergleichen die beste Passung aufwies (vgl. Anhang A08).

Dabei zeigte sich für keines der drei Kontextmerkmale ein signifikanter Haupteffekt auf die Itemschwierigkeit. Es zeigte sich ebenso kein signifikanter Haupteffekt, wenn alle drei Merkmale simultan in das Modell eingeführt wurden (vgl. Modell Kontext in Tabelle 24). Darüber hinaus konnten die Daten auch mit Interaktionen der Kontextmerkmale untereinander (vgl. Anhang A08)

sowie mit Interaktionen der Kontextmerkmale mit den Kompetenzniveaus (vgl. Anhang A09) nicht besser erklärt werden.

Allerdings hatte das Kompetenzniveau signifikanten Einfluss auf die Itemschwierigkeit: Items des mittleren ( $b=-0.31$ ,  $p<.05$ ) und des wissenschaftlichen Niveaus ( $b=-0.40$ ,  $p<.05$ ) waren signifikant schwieriger zu lösen als Items des unteren Kompetenzniveaus. Jedoch wies dieses Modell von den gerechneten Modellen die schlechteste Passung an die Daten auf ( $BIC=22736$ ) und konnte nur 7% der Itemschwierigkeitsvarianz aufklären.

**Abbildung 5 Mittlere geschätzte Schwierigkeiten der Items zu Verdunstung getrennt nach Inhalt**



*Anmerkungen.* falsche Zust., falsche Zustandsänderung; ad-hoc, Ad-hoc-Konstruktion; Nichterh., keine Massenerhaltung; wissenschaftsnah, wissenschaftsnahe Vorstellung;

Darüber hinaus beeinflusste der konkrete Inhalt der Items die Schwierigkeit bedeutsam (vgl. Abbildung 5). Alle Antwortalternativen, ausgenommen die korrekte Beschreibung der Zustandsänderung, waren signifikant schwerer als die fehlerhafte Beschreibung der Zustandsänderung. Dabei zeigt sich, dass die Ablehnung der anthropomorphen Deutung die höchste Herausforderung für die Probanden darstellte ( $b=-1.46$ ,  $p<.001$ ). Auch die Zustimmung zu der Beschreibung der Zustandsänderung ( $b=-1.26$ ,  $p<.001$ ) und zu der wissenschaftsnahen Antwortalternative ( $b=-0.78$ ,  $p<.001$ ) stellte sich als schwierig heraus. Leichter war die Ablehnung der ad-Hoc-Konstruktion bzw. der Nichterhaltungsvorstellung ( $b=-0.59$ ,  $p<.001$ ) zu bewältigen. Am einfachsten war die falsche Zustandsänderung ( $b=0.29$ ,  $p<.10$ ) und die korrekte Beschreibung des Ortswechsels ( $b=-0.12$ ,  $p=.36$ ) zu bearbeiten (siehe auch Abbildung 5). Von den untersuchten Modellen wies dieses die beste Passung auf die Daten auf ( $BIC=22611$ ) und klärte mit  $R^2=.58$  einen großen Teil der Itemschwierigkeitsvarianz auf.

#### 5.4.2.2 Kondensation

Im Bereich Kondensation zeigten die drei Kontextmerkmale Ort, Temperatur und Kondensationskern keinen Einfluss auf die Itemschwierigkeit (vgl. Tabelle 25). Allerdings fand sich ein signifikanter Haupteffekt des Kontextmerkmals Wasserquelle. Dabei waren Aufgaben ohne Quelle leichter zu lösen als solche mit ( $b=0.54$ ,  $p<.05$ ), wodurch bei guter Passung ( $BIC=4883$ ) eine Varianzaufklärung von  $R^2=.12$  erzielt wurde. Durch die weiteren Kontextmerkmale konnte jeweils nur ein kleiner Teil der Varianz aufgeklärt werden und auch die Passung der Modelle an die Daten war relativ schlechter (vgl. Tabelle 25). Darüber hinaus fand sich für keines der Kontextmerkmale eine signifikante Interaktion mit dem Kompetenzniveau (vgl. Anhang A10).

**Tabelle 25 Ergebnisse der Prüfung des Einflusses der Aufgabenmerkmale für Kondensation**

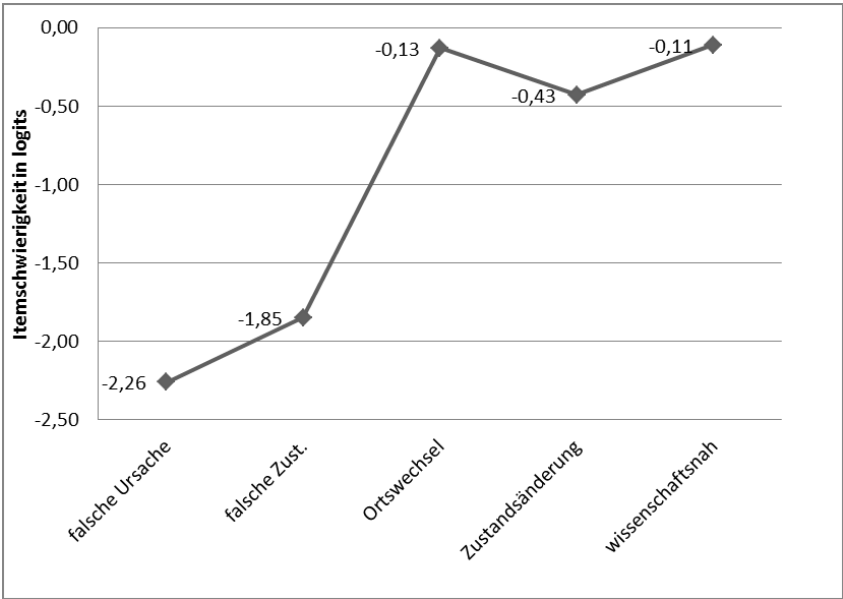
	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	R <sup>2</sup>
Nullmodell	-0.35	0.33	0.79	0.97	0.58	4913	
Niveaus			0.65	0.96	0.58	4911	.32
(NV)	-1.95***	0.48					
ZV	1.65***	0.39					
WV	1.83***	0.40					
Ort			0.79	0.97	0.58	4921	.00
(drinne)	-0.34	0.35					
draußen	-0.02	0.22					
Temperatur			0.79	0.97	0.58	4889	.01
(warm)	-0.42	0.35					
kalt	0.16	0.23					
Kondensationskern			0.79	0.97	0.58	4921	.00
Fläche	-0.36	0.34					
Dampf	0.04	0.26					
Quelle			0.74	0.97	0.58	4883	.12
(ja)	-0.58 <sup>+</sup>	0.34					
nein	0.54*	0.22					
Antwortalternative			0.64	0.96	0.58	4926	.35
(falsche Ursache)	-2.26**	0.80					
falsche Zust.	0.41	0.84					
Ortswechsel	2.13**	0.76					
Zustandsänderung	1.83*	0.75					
wissenschaftsnah	2.15**	0.75					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; falsche Zust., falsche Zustandsänderung; wissenschaftsnah, wissenschaftsnahe Vorstellung; <sup>+</sup> $p<.10$ ; \* $p<.05$ ; \*\* $p<.01$ ; \*\*\* $p<.001$ ;

Die Prädiktoren Kompetenzniveau ( $R^2=.32$ ) und Antwortalternative ( $R^2=.35$ ) wirkten beide bedeutsam auf die Itemschwierigkeit und konnten von den untersuchten Faktoren in diesem Bereich den größten Anteil an Varianz aufklären. Die Items auf dem mittleren ( $b=1.65$ ,  $p<.001$ ) und wissenschaftlichen ( $b=1.83$ ,  $p<.001$ ) Niveau waren dabei leichter zu lösen als Items des naiven Niveaus. Allerdings ist zu berücksichtigen, dass nur eine sehr geringe Anzahl von Antwortalternativen des naiven Niveaus in die Analysen einging.

Bei den verschiedenen Inhalten der Antwortalternativen war die Beschreibung einer falschen Ursache der Kondensation am schwierigsten zu lösen ( $b=-2.26$ ,  $p<0.01$ ). Die Beschreibung des Ortswechsels ( $b=2.13$ ,  $p<0.01$ ), die Beschreibung der Zustandsänderung ( $b=1.83$ ,  $p<0.05$ ) und die wissenschaftsnahe Vorstellung ( $b=2.15$ ,  $p<0.01$ ) waren signifikant leichter zu lösen (siehe Abbildung 6).

**Abbildung 6 Mittlere geschätzte Schwierigkeiten der Items zu Kondensation getrennt nach Inhalt**



Anmerkungen. falsche Zust., falsche Zustandsänderung; wissenschaftsnah, wissenschaftsnahe Vorstellung;

Trotz der höheren Varianzaufklärung zeigten sowohl das Modell mit dem Prädiktor Kompetenzniveaus ( $BIC=4911$ ) als auch mit dem Prädiktor Antwortalternative ( $BIC=4926$ ) eine schlechtere Anpassung an die Daten als das Modell mit dem Prädiktor Quelle ( $BIC=4883$ ).

#### 5.4.2.3 Dichte und Verdrängung

In den beiden Bereichen Dichte und Verdrängung konnte jeweils ein signifikanter Haupteffekt des Kontextmerkmals Kongruenz auf die Itemschwierigkeit belegt werden.

Items im Bereich Dichte, bei denen der Vergleich eines kongruenten Gegenstandspaares gefordert war, waren leichter zu lösen als inkongruente Items ( $b=3.06$ ,  $p<0.001$ ). Weder Items, in denen das Gegenstandspaar aus dem gleichen Material war ( $b=-0.40$ ,  $p=.42$ ), noch die Items, in denen das Gewicht konstant gehalten wurde ( $b=0.84$ ,  $p=.12$ ), unterschieden sich in der mittleren Schwierigkeit bedeutsam von inkongruenten Items. Der Faktor Kongruenz erklärte 91% der Schwierigkeitsvarianz (vgl. Tabelle 26).

**Tabelle 26 Ergebnisse der Prüfung des Einflusses der Aufgabenmerkmale für Dichte**

Feste Effekte		Zufällige Effekte			Modellkriterien	
B	SE	SD(item)	SD(person)	SD(class)	BIC	R <sup>2</sup>

Nullmodell	0.75	0.60	1.52	1.89	0.50	759	
Kongruenz			0.45	1.89	0.41	758	.91
(inkongruent)	-0.21	0.49					
gleiches Material	-0.40	0.49					
gleiches Gewicht	0.84	0.54					
kongruent	3.06***	0.52					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001;

Auch im Bereich Verdrängung waren die kongruenten Items leichter zu lösen als inkongruente ( $b=3.86$ ,  $p<.001$ ), während sich Items mit gleichem Gewicht nicht von den inkongruenten unterschieden ( $b=0.29$ ,  $p=.38$ ). Durch diesen Prädiktor konnte so viel Varianz aufgeklärt werden, dass keine Restvarianz der Itemschwierigkeit festzustellen war ( $R^2=1.00$ ; vgl. Tabelle 27).

**Tabelle 27 Ergebnisse der Prüfung des Einflusses der Aufgabenmerkmale für Verdrängung**

	Feste Effekte		Zufällige Effekte			Modellkriterien	
	B	SE	SD(item)	SD(person)	SD(class)	BIC	R <sup>2</sup>
Nullmodell	2.30***	0.65	1.28	2.09	0.12	354	
Kongruenz			0.00	0.00	0.00	350	1.00
(inkongruent)	1.52***	0.35					
gleiches Gewicht	0.29	0.33					
kongruent	3.86***	0.94					

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; <sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001;

## 5.5 Diskussion von Studie 2

### 5.5.1 Zusammenfassung und Diskussion der Ergebnisse von Studie 2

In Studie 1 dieser Arbeit konnte kein Hinweis auf eine starke Kontextualisierung der Vorstellungen von Grundschulern zu den Themenbereichen Schwimmen/Sinken und Verdunstung/Kondensation gefunden werden, die sich in Schwierigkeiten bei der Bearbeitung von Testaufgaben aus dem Projekt Science-P ausdrückte. Allerdings waren entsprechende Kontextmerkmale bei der Entwicklung der Aufgaben für diese Studie nicht berücksichtigt worden. Sie waren somit nicht gleichmäßig über die Items verteilt und zeigten zum Teil deutliche Abhängigkeiten untereinander (siehe Kapitel 4.5.4).

Daher wurden in der zweiten Studie dieser Arbeit die in Frage stehenden situativen Merkmale systematisch bei der Aufgabenkonstruktion variiert, während weitere Merkmale, wie z.B. formale Attribute der Aufgaben und der Inhalt der angebotenen Antwortalternativen, konstant gehalten wurden (siehe Kapitel 5.3.1.2). Die methodische Modifikation der Aufgabenkontexte stellte allerdings besonders für den Bereich Kondensation eine große Herausforderung dar, so dass nicht für alle möglichen Kombinationen der Kontextfaktoren Aufgabenstämme realisiert werden konnten. Darüber hinaus wurden bei der Selektion der Items anhand der eindimensionalen Rasch-

Skalierung lediglich die psychometrischen Eigenschaften der Items berücksichtigt, so dass in den Bereichen Verdunstung und Kondensation bestimmte Inhalte besonders stark von einem Ausschluss aus den weiteren Analysen betroffen waren. Dies bezieht sich bei Verdunstung besonders auf Items, die eine anthropomorphe Deutung der Verdunstung als mögliche Erklärung anboten. Bei Kondensation mussten nahezu alle Items des naiven Niveaus aus dem endgültigen Datensatz eliminiert werden.

In Zusammenhang mit dem Ausschluss fast aller naiven Items im Bereich Kondensation ist außerdem die insgesamt ungenügende psychometrische Qualität dieser Skala zu nennen. Dies ist möglicherweise darauf zurück zu führen, dass das Prinzip der Kondensation i.d.R. erst von den Schülern nachvollzogen werden kann, wenn ein Verständnis des gasförmigen Zustandes bereits erlangt wurde (Johnson, 1998; Tytler, 2000). Bevor dieses Lernziel erreicht wurde, haben die Probanden im Grundschulalter, besonders in den unteren Jahrgangsstufen, u.U. noch keine eigene Erklärung dieses Phänomens entwickelt, und auch in anderen Studien konnten nur anekdotische Erzählungen statt Erklärungen beobachtet werden (z.B. Tytler, 2000). In diesem Fall hätten die Grundschüler noch keinen internen Maßstab, gegen den sie die angebotenen Erklärungsansätze beurteilen könnten, was vermehrtes Raten und damit schlechtere psychometrische Qualität bedingt haben könnte. Die Ergebnisse der Analysen des selektierten Itempools werden im Folgenden zusammengefasst und hinsichtlich ihrer Bedeutung für die in Kapitel 5.2 aufgeworfenen Fragestellungen diskutiert.

#### *5.5.1.1 F2: Effekt der Kontextmerkmale auf die Aufgabenschwierigkeiten in Studie 2*

Die erste Fragestellung aus Studie 2 bezog sich auf die Wirkung der untersuchten Kontextmerkmale auf die Itemschwierigkeit. Die Schwierigkeiten der Items zu Verdunstung, die in Studie 1 zumindest marginal durch den Ort der zu erklärenden Phänomene beeinflusst wurde, wurden in dieser Studie durch keines der Kontextmerkmale mitbestimmt. Allerdings ist zu beachten, dass die Lehrkräfte der teilnehmenden Schüler angaben, dass das Thema bereits im Unterricht behandelt wurde (siehe Kapitel 5.3.1.1). Da allerdings auch angenommen wird, dass die Kontextabhängigkeit von Vorstellungen mit zunehmender Angemessenheit der Vorstellungen abnimmt (z.B. Engel Clough & Driver, 1986), ist ungeklärt, ob die Ergebnisse in einer Stichprobe von Schülern ohne Vorkenntnisse stärker kontextualisiert ausgefallen wären.

Der signifikante Einfluss des Merkmals Wasserquelle auf die Kondensationsitems konnte in dieser Studie hingegen bestätigt werden. Es scheint also tatsächlich der Fall zu sein, dass der Effekt des Ortes in Studie 1 bei Verdunstung durch andere, konfundierte Merkmale, wie z.B. den Inhalt der Antwortalternativen hervorgerufen wurde. Der Effekt der Wasserquelle bei Kondensation hingegen scheint überlagert gewesen zu sein und hatte in Studie 2 ein deutlicheres Ausmaß. Aufgaben, in denen keine potentielle Wasserquelle sichtbar war, waren leichter zu lösen als solche, in denen eine

sichtbar war. Dies kann dadurch bedingt sein, dass das Vorhandensein einer Wasserquelle dazu führt, dass diese, statt der umgebenden Luft, in die Erklärungen mit einbezogen wird.

Für den Bereich Dichte wurde in dieser Studie das Merkmal Kongruenz systematisch umgesetzt und konnte so einen großen Beitrag zur Erklärung der Schwierigkeitsvarianz leisten. Darüber hinaus wurde die Kongruenz auf den Bereich Verdrängung erfolgreich übertragen und klärte hier die Varianz sogar vollständig auf. Dieser Befund deckt sich im Bereich Dichte mit den Ergebnissen aus Studie 1 und mit den Ergebnissen anderer Autoren (Fassoulopoulous et al., 2003; Smith et al., 1985). Dabei waren Aufgaben mit kongruentem Gegenstandspaar jeweils deutlich leichter zu lösen als Aufgaben, bei denen das Gegenstandspaar inkongruente oder gleiche Merkmale aufwies. Es existieren zwei theoretische Möglichkeiten, wie ein solches Antwortmuster zustande gekommen sein kann: Die Probanden der Studie konnten ihr elaboriertes Konzept, welches sie bei kongruenten Aufgaben anzuwenden verstanden, nicht auf die anderen Aufgaben übertragen oder ihre naive Vorstellung, bei der die Begriffe Dichte und Gewicht noch nicht differenziert sind, sind ausreichend, um dennoch kongruente Aufgaben schon lösen zu können, aber unzureichend, um auch andere Aufgaben erfolgreich zu bearbeiten. Aus den Ergebnissen der vorliegenden Studie lässt sich nicht schlussfolgern, welche Annahme die zutreffendere ist. In beiden Fällen wird aber die Bedeutung bzw. Notwendigkeit des Einsatzes von unterschiedlichen Aufgaben zur Erfassung der Kompetenz in diesen Bereichen ersichtlich. Diese Forderung lässt sich des Weiteren auch auf Lernaufgaben übertragen: Sollen zum Erlangen des Verständnisses von Materialdichte Gegenstände verglichen werden, sollte ebenso darauf geachtet werden, dass sowohl kongruente wie inkongruente Gegenstandspaare bzw. Gegenstandspaare, bei denen ein Merkmal konstant gehalten wird, einbezogen werden.

Erstaunlich ist bei den Aufgaben zu Dichte und Verdrängung der enorme Anteil an Varianz, der aufgeklärt werden kann. Bei der Beurteilung der Varianzaufklärung ist allerdings zu bedenken, dass die Aufgaben in dieser Studie in formalen Merkmalen einander so weit wie möglich angeglichen wurden, um den Einfluss der Kontextmerkmale möglichst frei von Störeinflüssen zu messen. Geht man von einer Aufgabenentwicklung aus, bei der nicht auf strikte Parallelität der Aufgaben geachtet wird, ist davon auszugehen, dass der Anteil erklärter Varianz abnehmen würde. Für den Bereich Verdrängung ähnelt die untersuchte naive Vorstellung inhaltlich der im Bereich Dichte. In beiden Fällen wird das absolute Gewicht stellvertretend für eine andere Variable (Volumen im Falle von Verdrängung bzw. Gewicht relativ zum Volumen im Falle von Dichte) von den Probanden betrachtet. Ein Unterschied zwischen den beiden Bereichen liegt allerdings darin, dass in Bezug auf Verdrängung ein Verständnis für das Volumen, bzw. altersgerecht umschrieben die Größe eines Gegenstandes, vorhanden ist, während im Falle der Dichte eine differenzierte Vorstellung des zu erlernenden Konzeptes zunächst noch neu erworben werden muss (vgl. z.B. Smith et al., 1985). Inkongruente Aufgaben als Lernaufgaben könnten im Falle von Verdrängung also insbesondere dazu genutzt werden, den Schülern Grenzen ihrer Vorstellung aufzuzeigen (Der

schwere Gegenstand verdrängt in bestimmten Fällen nicht mehr, sondern weniger Wasser als der leichtere Gegenstand.) und auf eine plausible Erklärung aufmerksam zu machen (Die Gegenstände unterscheiden sich nicht nur hinsichtlich des Gewichts, sondern auch hinsichtlich des Volumens bzw. der Größe.). Systematische Vergleiche von Gegenständen bzw. gezielte Variation der in Frage stehenden Faktoren könnten in einem experimentierenden Unterricht gezielt zur Prüfung der Vermutungen der Schüler eingesetzt werden.

#### *5.5.1.2 F3: Interaktion zwischen Kontextmerkmalen und Kompetenzniveaus*

Eine mögliche Wechselbeziehung zwischen den untersuchten Kontextmerkmalen und den Kompetenzniveaus der Aufgaben konnte in dieser Studie nur für die Bereiche Verdunstung und Kondensation untersucht werden. In keinem Falle konnte eine solche Interaktion zur Erklärung der Daten beitragen. Es waren also keine Effekte der Kontextmerkmale differenziert nach dem Grad der Angemessenheit der untersuchten Schülervorstellungen in dieser Studie zu beobachten. Vor allem im Bereich Verdunstung könnte dies evtl. auf die Sophistizierung des Wissens der Schüler zurück zu führen sein, die mit diesem Konzept bereits im Unterricht konfrontiert wurden. Im Bereich Kondensation allerdings ist nicht davon auszugehen, dass die Mehrheit der Schüler bereits Unterricht zu diesem Konzept erhalten hatte.

#### *5.5.1.3 FZ: Einfluss des konkreten Iteminhalts auf die Aufgabenschwierigkeiten in Studie 2*

Neben den Kontextmerkmalen wurde in Studie 2 ebenfalls der Einfluss der Kompetenzniveaus auf die Itemschwierigkeit in den Bereichen Verdunstung und Kondensation untersucht. In beiden Fällen waren signifikante Effekte festzustellen. Items zu Verdunstung, die dem naiven Niveau zuzuordnen waren, waren leichter zu lösen als Aufgaben des mittleren und wissenschaftlichen Niveaus. Dieser Befund entspricht damit den Annahmen des Kompetenzmodells aus Science-P. Nimmt man allerdings die Analyse des Einflusses des konkreten Iteminhaltes hinzu, relativiert sich dieser Befund. Die Untersuchung dieses Faktors ergab ebenfalls einen signifikanten Effekt, der einen erheblichen Anteil Varianz aufklären konnte. Dabei zeigte sich allerdings, dass sich die Schwierigkeiten der Inhalte der Antwortalternativen nicht nach ihrem zugewiesenen Niveau anordneten, sondern sich eine Anordnung unabhängig vom angenommenen Kompetenzniveau, bei der sowohl naive, wie auch mittlere und wissenschaftliche Vorstellungen zu den besonders schwer zu lösenden Items zählten, abbildete. Eine Schwierigkeitsanordnung entsprechend der Kompetenzniveaus ergibt sich erst aus der Betrachtung der mittleren Schwierigkeiten der Niveaus über konkrete Inhalte hinweg und kann somit auch die relativ geringe Varianzaufklärung durch die Kompetenzniveaus sowie die eher schlechte Passung dieses Modells erklären.

Im Bereich Kondensation kehrte sich der Effekt der Niveaus sogar um: Items mit naiven Vorstellungen waren signifikant schwieriger zu lösen als Items des mittleren bzw. wissenschaftlichen Niveaus. Der Effekt widersprach damit genau den Annahmen des zugrundeliegenden Kompetenzmodells. Die Varianzaufklärung, die durch diesen Faktor erreicht



wurde, war annähernd genau so groß wie die, die durch die Aufklärung mittels des konkreten Iteminhalts erzielt werden konnte. In dem Modell, das den Iteminhalt untersuchte, zeigte sich zudem, dass die Beschreibung einer falschen Ursache auf dem naiven Niveau besonders schwierig zu lösen war. Bei der Interpretation dieses Befundes ist zu beachten, dass bei naiven Vorstellungen die Ablehnung als richtige Lösung gewertet wurde. Im Falle von Kondensation, welches ein sehr anspruchsvolles Konzept darstellt und von den Schülern i.d.R. erst verstanden werden kann, nachdem eine Vorstellung des gasförmigen Zustandes erreicht wurde (Johnson, 1998; Tytler, 2000), scheinen sich naive Vorstellungen besonders hartnäckig und lange zu halten, während die Zustimmung zu vorgegebenen elaborierten Vorstellungen leichter gelingt. Des Weiteren ist allerdings zu bedenken, dass die meisten naiven Vorstellungen aufgrund mangelnder Trennschärfe ausgeschlossen wurden. Diese Schlussfolgerung fußt also auf einer schmalen empirischen Basis und sollte dringend weiteren Untersuchungen unterzogen werden.

Dass auch naive Vorstellungen mit hoher Wahrscheinlichkeit angenommen wurden, kann auch in der Vereinbarkeit der verschiedenen Antwortalternativen begründet sein: Die Beschreibung, dass Wasser sich in der Luft verteilt, schließt z.B. nicht zwangsläufig aus, dass noch etwas anderes mit dem Wasser passiert sein könnte. Die Erklärungen in den untersuchten Aufgaben bezogen sich nicht explizit auf „das ganze Wasser“ oder „alles Wasser, das jetzt weg ist“. Darüber hinaus sollten die Itemantworten in dem Fragebogen für das Projekt Science-P sich bewusst nicht gegenseitig logisch ausschließen, um mögliche Abhängigkeiten zu minimieren. Neben solch formalen Überlegungen deckt sich der Befund allerdings auch mit theoretischen Annahmen und empirischen Belegen anderer Autoren, dass häufig naive und fortgeschrittenere Vorstellungen parallel gehalten werden (vgl. z.B. Caravita, 2001; Vosniadou & Ioannides, 1998).

#### *5.5.1.4 Zusammenfassung*

Während in den Bereichen Dichte und Verdrängung das Kontextmerkmal Kongruenz den größten Anteil der Varianz aufklären konnte, erklärte in den Bereichen Verdunstung und Kondensation also der konkrete Iteminhalt am stärksten das Antwortverhalten der Probanden. Dabei überlagerte der konkrete Inhalt auch Effekte eines allgemeineren Niveaus der Vorstellungen, d.h. auch Vorstellungen, die aus fachlicher Perspektive ungenügend waren, wiesen hohe Schwierigkeiten auf.

#### *5.5.2 Beschränkungen der Studie 2*

Für Studie 2 wurden Items aus Studie 1 ausgewählt und einer umfassenden Überarbeitung unterzogen, um eine Untersuchung der Kontextmerkmale frei von dem Einfluss weiterer konfundierter Merkmale der Aufgaben zu ermöglichen. Darüber hinaus war es teilweise notwendig neue Aufgaben zu entwickeln, um das experimentelle Design in Studie 2 konsequent umzusetzen. Dieses Ziel der Aufgabenentwicklung brachte einige Einschränkungen mit sich: Die Auswahl möglicher Erklärungsansätze für naturwissenschaftliche Phänomene, die als Antwortalternativen

angeboten wurden, wurde v.a. in den Bereichen Verdunstung und Kondensation deutlich eingeschränkt. Das Aufgabenformat der Querschnittstudie wurde in einer vorhergehenden Studie mit zufriedenstellenden Ergebnissen auf seine Validität überprüft (vgl. Pollmeier et al., 2011; siehe auch Kapitel 4.3.1.1). Es ist allerdings eine offene Frage, ob durch die Veränderung der Aufgaben dieser Befund uneingeschränkt auf die Aufgaben der Experimentalstudie übertragbar ist. Es wurden zwar v.a. häufig beobachtete Schülervorstellungen als Erklärungen berücksichtigt, jedoch war in den durchgeführten Interviews die Vielfältigkeit der Vorstellungen ein zentrales Merkmal und diese Vielfältigkeit konnte durch den Fragebogen nicht mehr abgebildet werden. Die Fragebogenergebnisse bzw. Antworten der Schüler sind daher eher als Zustimmung zu gängigen Vorstellungen zu interpretieren.

Darüber hinaus war die Experimentalstudie im Rahmen des Projektes Science-P als Pilotierungsstudie angelegt und die Stichprobengröße für diesen Zweck im Vergleich zur Querschnittstudie deutlich geringer. Die Untersuchung ist daher eher als explorative Analyse im Hinblick auf die Aufgabenmerkmale zu verstehen. Um verlässlichere Aussagen zu erhalten, sollten die Ergebnisse an einer größeren, möglichst repräsentativen Stichprobe von Grundschulkindern abgesichert werden.

In Studie 2 wurden, vor den Analysen der Daten hinsichtlich der Kontextmerkmale, die Items wie auch in Studie 1 hinsichtlich ihrer psychometrischen Qualität untersucht und selektiert (siehe Kapitel 5.3.2.3). Dabei wurden in den Bereichen Verdunstung und Kondensation besonders viele Items des naiven Niveaus selektiert. Des Weiteren wurde die ungenügende psychometrische Qualität im Bereich Kondensation aufgedeckt. Obwohl diese Probleme mit der Güte der Testaufgaben in Studie 1 nicht auftraten und unklar ist, in welchem Maße sie in dieser Studie auch durch die geringe Stichprobengröße bedingt waren, sollte der Verbesserung der Testaufgaben weiter Aufmerksamkeit geschenkt werden.

### **5.5.3 Konsequenzen für die Aufgabenentwicklung aus Studie 2**

Wie in Kapitel 4.5.3 gefordert, wurden in dieser Studie die Kontextmerkmale als wichtiges Kriterium in der Aufgabenentwicklung berücksichtigt, während gleichzeitig andere Merkmale eingeschränkt bzw. kontrolliert berücksichtigt wurden. Wie schon in Kapitel 5.5.2 erörtert, war dies zwar die optimale Vorgehensweise für die Untersuchung der hier beschriebenen Fragestellungen, könnte allerdings zu einer eingeschränkten Validität des Tests geführt haben. Daher bleibt die Forderung nach einer bewussten Auswahl der Aufgabenkontexte, die auch die Perspektive der angestrebten Probandengruppe neben fachlichen Gesichtspunkten berücksichtigen sollte, bestehen. Zusätzlich sollte der konkrete Inhalt angebotener Antwortalternativen, neben möglichen globalen Kompetenzniveaus, Beachtung finden. Dazu sollte v.a. berücksichtigt werden, welche Antworten bzw. eigene Erklärungen eine vergleichbare Gruppe tatsächlich auf die untersuchten Phänomene

liefert und welche empirischen Schwierigkeiten die so gewonnenen Antwortalternativen in einem schriftlichen, geschlossenen Antwortformat erzielen.

Für die Bereiche Verdunstung und Kondensation allerdings tritt die Forderung der Berücksichtigung des Kontexts aufgrund der schwachen Befundlage zur Kontextualität des Antwortverhaltens in den Hintergrund. Da allerdings nicht auszuschließen ist, dass andere als die hier untersuchten situativen Merkmale das Antwortverhalten von Grundschulern beeinflussen könnten, sollte die Selektion geeigneter Aufgabenkontexte zumindest im Bewusstsein für potentielle Effekte geschehen.

Für die Bereich Dichte und Verdrängung sollte, aufgrund der massiven Varianzaufklärung, die in dieser Studie erreicht werden konnte, das Merkmal Kongruenz bei der Auswahl von Aufgabenkontexten unbedingt berücksichtigt werden. Dabei sollten zu einer vollständigen Erfassung des Wissens in diesem Bereich alle möglichen Aufgabentypen (kongruent, inkongruent, eine Dimension konstant gehalten) berücksichtigt werden. Die Übertragung dieses Merkmals auf andere Aufgabenformate (z.B. Vergleich der Dichte von Vollkörpern mit der Dichte von Wasser zur Vorhersage des Schwimmverhaltens) ist möglich. Obwohl zwar für eine solche Ausweitung noch keine empirischen Befunde vorliegen, legen die Ergebnisse nahe, dass die Kongruenz auch in weiteren Bereichen, in denen das Verständnis von Dichte und Verdrängung eine Rolle spielt, ein wichtiger Einflussfaktor sein könnte.

#### **5.5.4 *Fazit aus Studie 2 und Ausblick***

Die zentrale Frage dieser Arbeit nach einem möglichen systematischen Einfluss von Kontextmerkmalen auf das Antwortverhalten von Grundschulern kann auf Basis der Ergebnisse aus Studie 2 differenziert nach dem Themenbereich beantwortet werden: Während in den Bereichen Verdunstung und Kondensation keine Schwierigkeitsunterschiede im Hinblick auf situative Faktoren zu beobachten waren, sondern der konkrete Iteminhalt der wichtigste erklärende Faktor war, konnte in den Bereichen Dichte und Verdrängung mit der Kongruenz ein zentrales schwierigkeiterzeugendes Merkmal identifiziert werden. Darüber hinaus ist festzuhalten, dass auch in den Bereichen Verdunstung und Kondensation im Hinblick auf Kompetenzniveaus keine Kohärenz des Antwortverhaltens zu finden war, sondern vielmehr naive und fortgeschrittene Erklärungen simultan zur Erklärung naturwissenschaftlicher Phänomene gewählt wurden.

Zur weiteren Untersuchung der Fragestellungen dieser Arbeit wurde das Merkmal Kongruenz in der nachfolgenden Längsschnittstudie innerhalb des Projektes Science-P eingesetzt. Dabei wurde die Nutzung des Aufgabenmerkmals auf weitere Aufgabentypen aus den Bereichen Dichte und Verdrängung übertragen. Die Auswertungen der Daten dieser Studie sind noch nicht abgeschlossen.

In früheren Studien wurde zwar häufig eine Kontextualität von Schülervorstellungen berichtet (vgl. z.B. Tytler, 2000; siehe Kapitel 2.2.4.2 und 2.2.5.2). Allerdings waren diese Befunde häufig

Nebenprodukte von Untersuchungen, die eine andere Zielsetzung hatten und die Konstruktion der Aufgaben erlaubte lediglich den Vergleich sehr unterschiedlicher Kontexte bzw. die Aufgabenkontexte unterschieden sich hinsichtlich mehrerer konfundierter Merkmale. Auch wurde häufig ein offenes Interviewformat statt eines Fragebogens eingesetzt. Vor allem für Bereiche, in denen dies noch nicht systematisch untersucht wurde, wäre es also interessant, die Befunde dieser Arbeit auf ein offenes, mündliches Antwortformat zu übertragen.

## 6 Fazit

### 6.1 Zusammenfassung und Diskussion der beiden Studien

Zusammenfassend wiesen beide Studien dieser Arbeit recht unterschiedliche Befunde auf: In Studie 1 waren im Themenbereich Verdunstung/Kondensation tendenziell Hinweise auf eine Kontextualisierung des Wissens zu finden. Im Themenbereich Schwimmen/Sinken hingegen zeigte sich nur für den Bereich Dichte eine Abhängigkeit der Itemschwierigkeiten von einem Kontextmerkmal. Zudem konnten in Studie 1 sehr schwache Hinweise für eine zunehmende Dekontextualisierung der Vorstellungen bei fortschreitendem Niveau des Wissens finden.

In Studie 2 konnten die Hinweise auf eine Kontextualisierung des Wissens im Themenbereich Verdunstung/Kondensation nicht etwa repliziert oder sogar ausgebaut werden, sondern die untersuchten Kontextmerkmale zeigten (bis auf eines im Bereich Kondensation) keinerlei Effekt auf die Itemschwierigkeiten. Eine abnehmende Kontextualisierung war ebenfalls nicht festzustellen. Hingegen konnte der konkrete Iteminhalt als bedeutsamer Prädiktor des Antwortverhaltens identifiziert werden. In den beiden Bereiche Dichte und Verdrängung konnte das Merkmal Kongruenz nahezu die gesamte Schwierigkeitsvarianz der Items aufklären. Eine zunehmende Dekontextualisierung ließ sich für diese beiden Bereiche in Studie 2 nicht untersuchen.

Als zentrale Bedingung für die themenspezifisch divergent ausgeprägte Kontextualisierung der Schülervorstellungen wurde in Kapitel 4.5.1.5 die Unterschiedlichkeit der beiden Themenbereiche diskutiert. Als wesentliche Diskrepanz zwischen den Themenbereichen wurde dabei eine größere Vielfältigkeit von naiven Schülervorstellungen sowie eine größere Anzahl untersuchter fachlicher Aspekte bei Schwimmen/Sinken beschrieben. Dieser Unterschied war in Studie 2 nicht mehr gegeben. Im Themenbereich Schwimmen/Sinken wurden zwei klar umgrenzte naive Vorstellungen (keine Differenzierung von Gewicht und Dichte, Verdrängung hängt von dem Gewicht ab) zwei ebenso klar umgrenzten wissenschaftlichen Vorstellungen gegenüber gestellt (proportionales Verständnis der Dichte, Verdrängung hängt vom Volumen ab). Auch im Themenbereich Verdunstung/Kondensation wurde statt vieler unterschiedlicher Vorstellungen nur eine Auswahl von je sieben Vorstellungen der verschiedenen Niveaus zur Auswahl gestellt.

Diese Veränderung an der Operationalisierung der Themenbereiche könnte erklärend für die unterschiedlichen Ergebnisse aus Studie 1 und 2 sein. Gleichzeitig kann die Veränderung zwischen den beiden Studien auch als eine Fokussierung auf einen kleineren, umgrenzten Ausschnitt aus beiden Themenbereichen betrachtet werden: Im Themenbereich Verdunstung/Kondensation wird nur eine begrenzte Auswahl möglicher Erklärungsansätze bzw. Schülervorstellungen betrachtet. Bei Schwimmen/Sinken wird eine fachliche Auswahl von Inhalten getroffen. Diese zweite Sichtweise kann erklären, wieso die wenigen gefundenen Effekte in Studie 2 größer ausfallen. Zudem verdeutlicht diese Betrachtungsweise die Relevanz einer genau durchdachten

Vorgehensweise zur Entwicklung valider Messinstrumente: Je nach Entscheidung für einen umfassenderen oder kleineren Ausschnitt aus dem zu messenden Konstrukt, treten sehr unterschiedliche Befunde zu Tage.

Die Unabhängigkeit der Schülervorstellungen vom Kontext im Themenbereich Verdunstung/Kondensation, vor allem in Studie 2, steht im Gegensatz zu den in den Kapitel 2.2.4.2 berichteten Befunden zur Kontextualität aus bestehender Forschung. In den berichteten Studien wurde allerdings ausnahmslos ein offenes Antwortformat eingesetzt, mehrheitlich mit einer mündlichen Befragung, während die Probanden in meiner Arbeit geschlossene schriftliche Aufgaben bearbeiteten. Die unterschiedlichen Anforderungen, die diese beiden Aufgabenmodi an die Probanden stellen, können möglicherweise den Einfluss des Aufgabenkontexts moderieren. Wie in Kapitel 2.2.3 erläutert, wird der Einfluss des Kontexts auf geäußerte Schülervorstellungen durch assoziative Prozesse mit zuvor gemachten Erfahrungen erklärt. Dabei wird die aktuelle Situation, für die z.B. Erklärungen geäußert werden sollen, mit Hilfe bereits bekannter Situationen interpretiert, die ihrerseits bestimmte situationsspezifische Merkmale aufwiesen. Dieser Prozess tritt bei Aufgaben bzw. Situationen mit bereits vorgegebenen Erklärungsansätzen, wie bei geschlossenen Antwortformaten, möglicherweise in den Hintergrund. Statt zunächst das erläuterte Phänomen zu betrachten, fokussieren die Probanden möglicherweise vermehrt auf die Antworten und bewerten diese hinsichtlich ihrer Plausibilität ohne dabei auf bereits gemachte Erfahrungen zurück greifen zu müssen. Die Merkmale von zuvor erlebten Situationen können somit das Antwortverhalten nicht beeinflussen. Dass das Kontextmerkmal Kongruenz dennoch einen Einfluss auf die geschlossenen, schriftlichen Aufgaben zu Dichte und Verdrängung hatte, ist wiederum durch die veränderten Aufgabenanforderungen erklärbar. Während bei Verdunstung und Kondensation Erklärungen für die beschriebenen Phänomene gefunden bzw. bewertet werden sollen, müssen die Probanden in diesen Bereichen basierend auf den gegebenen Informationen den Gegenstand mit einer bestimmten Eigenschaft auswählen, ohne ihre Entscheidung zu begründen bzw. zu erklären.

Diese zuletzt genannten Befunde meiner Arbeit zum Merkmal Kongruenz zeigten sich, insbesondere im Bereich Dichte, stabil über beide Studien hinweg. Darüber hinaus waren die statistischen Effekte durch dieses Merkmal in beiden Bereichen Dichte und Verdrängung beachtlich. Dieses Ergebnis kann daher als besonders bedeutsam erachtet werden. In den nächsten beiden Kapiteln werde ich daher darauf eingehen, welche Nutzen die Befunde meiner Arbeit in Bezug auf die Aufgabenentwicklung (siehe Kapitel 6.2) und Kompetenzmodellierung (siehe Kapitel 6.3) haben und dabei insbesondere auf die Ergebnisse in den Bereichen Dichte und Verdrängung eingehen.

## 6.2 Konsequenzen für die Aufgabenentwicklung

Wie die Ergebnisse der Aufgabenanalyse in Studie 1 (siehe Kapitel 4.4.1) zeigten, waren die Kontextmerkmale in den Aufgaben teilweise sehr extrem verteilt. Dies veranlasst zu der Empfehlung während der Aufgabenkonstruktion, auch bei Studien, die nicht den Einfluss situativer Merkmale untersuchen wollen, eine bewusste Auswahl der Anwendungskontexte vorzunehmen. Die Selektion der Kontexte kann die Operationalisierung des Konstruktes beeinflussen, da sie für die Untersuchung von Kompetenzen einerseits einen Teil der Konstruktdefinition darstellt (Klieme & Leutner, 2006; Weinert, 2002) bzw. bei unbalancierter Auswahl zu einer Messung von Kompetenzen in sehr speziellen, nicht repräsentativen Situationen führt. Diese Forderung erhält weitere Relevanz, wenn es um die Erfassung von Schülervorstellungen geht, da auch für diesen Bereich die Bedeutung des Kontextes bereits mehrfach beschrieben wurde (vgl. Kapitel 2.2.3). Eine Analyse der Kontexte sollte also entweder zusätzlich zur oder als der Teil Operationalisierung des Konstruktes vorgenommen werden. Leitende Motive können neben fachwissenschaftlichen Abwägungen dabei v.a. die Erfahrungswelt der untersuchten Zielgruppe und die natürliche Auftretenshäufigkeit verschiedener Kontextfaktoren sein. Im Optimalfall könnten der Aufgabenentwicklung Untersuchungen vorangestellt werden, in denen Situationen aus der Erfahrungswelt der Probanden hinsichtlich ihrer zentralen Merkmale eingeschätzt werden. Erste Ansätze dazu liefert z.B. das Paradigma der *typically perceived situations* (Joung, 2009; vgl. Kapitel 2.2.5.1). Eine systematische Variation von Kontexten oder Kontextmerkmalen könnte auf solchen Untersuchungen aufbauen, ist aber nur dann unbedingt erstrebenswert, wenn der Einfluss der Faktoren auch tatsächlich statistisch erfasst werden soll und sollte dann begleitet werden von einer Kontrolle möglicher weiterer Einflussfaktoren (vgl. Kapitel 4.5.2).

Neben den Kontexten sollte dem konkreten Inhalt angebotener Antwortalternativen die größte Beachtung geschenkt werden. Sowohl die Ergebnisse dieser Arbeit wie auch Befunde weiterer Analysen im Projekt Science-P (vgl. Pollmeier, Tröbst & Möller, 2012) belegen den großen Einfluss, den diese auf das Antwortverhalten der Probanden haben.

Sehr konkrete Hinweise für die Aufgabenentwicklung liefert meine Arbeit v.a. für die Bereiche Dichte und Verdrängung. Mit Hilfe dieses Merkmals können gezielt Aufgaben unterschiedlicher Schwierigkeit entwickelt werden. Die Beachtung dieses Merkmals bei der Aufgabenentwicklung kann daher einerseits dazu genutzt werden den eingesetzten Aufgabenpool zur Erfassung der Kompetenz in den Bereichen Dichte und Verdrängung an das Kompetenzniveau der untersuchten Probandengruppe anzupassen, d.h. bei besonders fähigen Personengruppen sollten vermehrt Aufgaben mit höherer Schwierigkeit bzw. inkongruente Aufgaben eingesetzt werden, um möglichst trennscharf zwischen Personen ähnlichen Kompetenzniveaus differenzieren zu können. Andererseits sollte eine vollständige Erfassung der Kompetenz in diesen Bereichen alle Aufgabentypen (kongruent, gleiche Dichte/gleiches Gewicht, gleiches Material, inkongruent) berücksichtigen. Dabei kann das Merkmal auf weitere Aufgabenarten übertragen werden. Zur

Erfassung der Kompetenz im Themenbereich Schwimmen/Sinken sollen Probanden z.B. Würfel unterschiedlicher Dichte mit einem imaginären Wasserwürfel vergleichen (vgl. z.B. Engelen et al., 2002; Hardy et al., 2006). Auch die dabei angestellten Vergleiche können kongruent oder inkongruent sein. Die Kongruenz könnte auch in diesem Fall einen entscheidenden Einfluss auf die Aufgabenschwierigkeit haben. Die Auswertungen entsprechender Daten einer Längsschnittstudie des Projektes Science-P stehen noch aus.

### **6.3 Konsequenzen für die Kompetenzmodellierung**

Wie in Kapitel 2.1.2.5 beschrieben wurde, besteht für die Modellierung von Kompetenzen v.a. im Bereich naturwissenschaftlicher Bildung im Grundschulalter noch erheblicher Forschungsbedarf. Eine besondere Herausforderung liegt dabei in der Entwicklung valider Instrumente zur Erfassung dieser Kompetenzen. Ein Rückgriff auf Befunde aus der Conceptual Change-Forschung kann dafür wertvolle Hinweise liefern. Insbesondere Erkenntnisse zum Einfluss des Kontextes auf Schülervorstellungen können die Auswahl geeigneter Aufgabenkontexte leiten und darüber hinaus eine Möglichkeit zur Beschreibung von Kompetenzentwicklung mittels fortschreitender Dekontextualisierung, d.h. elaborierte Erklärungsansätze können, unabhängig von Merkmalen der Anwendungssituation, stabil angewendet werden, liefern. Eine solche Modellierung würde eine Betrachtung des Antwortverhaltens über mehrere Aufgaben hinweg erfordern. Dabei könnten Kontextmerkmale Hinweise zur Zusammenstellung geeigneter Aufgabengruppen bieten. Konkrete Hinweise dazu lassen sich aufgrund der unklaren Befundlage aus der vorliegenden Arbeit allerdings nur schwer ableiten.

Neben der Dekontextualisierung wäre eine Beschreibung der zunehmenden Kohärenz der Angemessenheit der Vorstellungen eine Möglichkeit der Beschreibung der Kompetenzentwicklung. Die Ergebnisse der Studie 2 lassen den Schluss zu, dass naive und elaborierte Vorstellungen von den Probanden der Studie in den Bereichen Verdunstung und Kondensation parallel gehalten werden. Eine höhere Kompetenz könnte sich durch eine konsequente Ablehnung naiver Vorstellungen, bei gleichzeitiger Zustimmung zu fortgeschrittenen Vorstellungen, auszeichnen. Dazu wäre eine simultane Bepunktung von Items verschiedener Kompetenzniveaus notwendig.

Generell sollte die Dimension des Kontextes stärker einbezogen und konkretisiert werden (siehe Kapitel 6.2). Ausgehend von der definitorischen Bedeutung der Anwendungssituationen (Klieme & Leutner, 2006; Weinert, 2002) liegt auch hier das Potential eine Entwicklung der Kompetenz durch deren Ausweitung auf eine breitere Auswahl von Kontexten zu modellieren. Auf Grundlage der Befunde der vorliegenden Arbeit wird dies, wie schon beschrieben, für die Bereiche Dichte und Verdrängung empfohlen. Mit Hilfe des Merkmals Kongruenz können für die Aufgaben dieser Bereiche aufbauend auf Aufgabenanforderungen konkret unterschiedliche Kompetenzniveaus



beschrieben werden. Zudem bietet die Betrachtung von Antwortmuster über Aufgaben unterschiedlicher Ausprägung der Kongruenz hinweg u.U. die Möglichkeit Subgruppen von Schülern mit unterschiedlichen Verständnisstufen zu identifizierten (vgl. Pollmeier, 2014; Smith, 2007).

#### **6.4 Bedeutung der Ergebnisse der Studien 1 und 2 für die Debatte um Kohärenz vs. Fragmentierung**

In Kapitel 2.2.2 wurde dargestellt, dass die Frage nach der Beschaffenheit der mentalen Organisation und Struktur des anfänglichen naturwissenschaftlichen Wissens noch eine große offene Frage innerhalb der Conceptual Change-Forschung darstellt und dass unterschiedliche Ansätze zur Beantwortung dieser Frage die Kohärenz oder die Fragmentierung der Schülervorstellungen betonen. Zur empirischen Klärung dieser Problemstellung werden Studien heran gezogen, in denen die Konsistenz der Antwortprofile der Probanden über mehrere Aufgabenkontexte hinweg untersucht wird. Daher soll in diesem Kapitel untersucht werden, ob die Ergebnisse meiner Arbeit einen Beitrag zur Lösung dieser Frage leisten können.

Die Ergebnisse von Studie 1 weisen nur tendenziell auf ein eher lose organisiertes anfängliches Wissen, welches sich an situativen Merkmalen der Aufgaben orientiert, so wie es im Ansatz von diSessa (2008) beschrieben wird. Neben der kompartimentalisierten Struktur sagt der Fragmentierungsansatz eine Abnahme der Kontextualisierung mit zunehmender Kompetenz vorher (z.B. diSessa, 2006; diSessa & Sherin, 1998). Obwohl in Studie 1 für zwei Kontextmerkmale eine entsprechende signifikante Interaktion zu den postulierten Kompetenzniveaus gefunden wurde, ist den Daten eine solche Tendenz im Allgemeinen nicht zu entnehmen. Eine Entscheidung zugunsten des Fragmentierungs- oder des Kohärenzansatzes aufgrund der Ergebnisse der Querschnittstudie zu treffen, ist wegen der uneinheitlichen Befundlage schwierig.

In Studie 2 zeigte im Themenbereich Verdunstung/Kondensation die Betrachtung der Schwierigkeiten der verschiedenen Antwortalternativen, die den konkreten Inhalt der Items bildeten, dass v.a. das Ablehnen naiver Vorstellungen besondere Herausforderungen an die Probanden der Studie stellt und dass naive und fortgeschrittene Vorstellungen parallel gehalten werden. Dieses Phänomen erinnert an die von Vosniadou und Kollegen (Vosniadou, 2012; Vosniadou et al., 2008) beschriebenen synthetischen Modelle, in denen eigene Präkonzepte sowie wissenschaftliche Konzepte aus dem Unterricht von Schülern integriert werden, bevor ein vollständig wissenschaftliches Verständnis erreicht werden kann.

Die Fragmentierung, d.h. Kontextabhängigkeit, der Schülervorstellungen, die in Studie 2 für den Themenbereich Schwimmen/Sinken, stellvertretend untersucht an den beiden Konzepten Dichte und Verdrängung, festgestellt werden konnte, zeigt, dass die Schüler im Fall von kongruenten Aufgaben auch mit einer fehlerhaften Vorstellungen zu richtigen Lösungen der Aufgaben gelangen

können (vgl. Kapitel 5.5.1.1). In diesem Fall werden ihre naiven Vorstellungen weder Außenstehenden (z.B. Lehrpersonen) offenbar, noch führen falsche Vorhersagen zu einer Unzufriedenheit mit den eigenen Konzepten und somit zu der Motivation sich geeignetere Vorstellungen und Erklärungsansätze anzueignen. Wie von diSessa und Kollegen (diSessa, 2008; diSessa et al., 2004) beschrieben, kann erst bei fortschreitender Elaboriertheit eine Dekontextualisierung des Verständnisses stattfinden, die eine konstante Lösung der Aufgaben zu Dichte und Verdrängung ermöglicht. Allerdings können die Ergebnisse in diesem Bereich auch im Sinne des Kohärenzansatzes interpretiert werden: Eine naive Theorie von Dichte bzw. Verdrängung, bei der das Gewicht als zentrale Variable konzeptualisiert wird, führt zu einem inkohärenten Antwortmuster, bei dem kongruente Aufgaben gelöst werden können, während diese Strategie bei den anderen Aufgabentypen zu inkorrekten Antworten führt. Wird diese naive Theorie von einer elaborierteren abgelöst, zeigt sich ein kohärentes Antwortmuster, dass bei allen Aufgabentypen in den Bereichen Dichte und Verdrängung zu korrekten Antworten führt. Eine Überlegenheit eines der beiden Ansätze (Kohärenz vs. Fragmentierung) ist für diesen Bereich nicht zu erkennen. Also ist auch nach Studie 2 wieder keine übergreifende Schlussfolgerung möglich.

Die leichten Hinweise auf Kontextabhängigkeit der Vorstellungen, die in Studie 1 v.a. für Verdunstung und Kondensation, und in Studie 2 für Dichte und Verdrängung gefunden wurden, weisen auf eine kontextspezifische Aktivierung von Vorstellungen wie sie von Vertretern des Fragmentierungsansatzes beschrieben werden (diSessa, 2006; diSessa et al., 2004). Demnach werden einzelne informationshaltige Wissens Elemente je nach Situation durch Assoziation aktiviert und zur Erklärung der Situation bzw. des Phänomens genutzt (diSessa, 2006). Einschränkend ist allerdings anzumerken, dass diese Kontextabhängigkeit sich weder stabil – über Themenbereiche und Studien hinweg - zeigte, noch dass sie für elaboriertere Vorstellungen weniger ausgeprägt vorzufinden war. Dies könnte allerdings z.T. auf das Antwortformat zurückzuführen sein: Während häufig Interviewstudien Belege für die Kontextualität von Schülervorstellungen lieferten (z.B. diSessa et al., 2004; Tytler, 2000), wurde in der vorliegenden Arbeit bewusst, v.a. wegen des Bezuges zur Kompetenzmodellierung, ein geschlossenes schriftliches Antwortformat gewählt. Aus anderen Studien ist allerdings bekannt, dass v.a. im Grundschulalter mündliche und schriftliche bzw. offene und geschlossene Formate zu unterschiedlichen Ergebnissen führen können (z.B. Bullock et al., 2009; Pollmeier et al., 2011). Da bei vorgegebenen Erklärungsansätzen u.U. assoziative Vorgänge bei der Verarbeitung des Aufgabenstammes konkurrieren mit der Verarbeitung der Antwortalternativen, werden möglicherweise keine oder weniger starke Verknüpfungen zu eigenen Vorerfahrungen, die eine wichtige Quelle vorunterrichtlicher Vorstellungen darstellen (Driver & Easley, 1978; diSessa, 2006; Engel Clough & Driver, 1986), aktiviert und eine Kontextualität wird nicht erkennbar.

Dass keine generellen, stabilen Indikatoren für Kontextabhängigkeit der Schülervorstellungen gefunden wurde, kann auch als Hinweis auf eine kohärente, theorieähnliche Struktur des Wissens

gewertet werden (Vosniadou, 1994; Vosniadou et al., 2008). Den Probanden der Studien gelingt es, unabhängig von den untersuchten situativen Merkmalen, einen eigenen Erklärungsansatz zu finden und diesen kohärent auf die verschiedenen Kontexte anzuwenden. Allerdings kann auch diese Interpretation, mit Hinblick auf die gefundene Kontextabhängigkeit in den Bereichen Dichte und Verdrängung, vorerst nur eingeschränkt gelten. In diesem Fall ist auch für diese Befunde eine Interpretation im Sinne zweier sich ablösender Theorien möglich. Im Falle der gefundenen kontextuellen Abhängigkeit im Themenbereich Verdunstung/Kondensation in Studie 1 ist eine solche alternative Interpretation allerdings schwieriger möglich. Eine Vermutung ist auch hier ein möglicher Einfluss des Antwortformates: Durch die Vorgabe einer größeren Anzahl verschiedener naiver Vorstellungen (im Vergleich zu Studie 2) werden die Probanden evtl. dazu verleitet, weniger konsistent zu antworten. In Studie 2 zeigte sich hingegen, dass z.T. naive und elaboriertere Vorstellungen parallel gehalten werden. Ähnlich wird die Integration naiver und wissenschaftlicher Erklärungen in synthetischen Modellen von Vosniadou und Kollegen (2008) beschrieben. Die Ursache dieser synthetischen Modelle liegt laut den Autoren (Vosniadou, 2012; Vosniadou et al., 2008) v.a. in einem unangemessenem Umgang mit dem Vorwissen der Schüler im Unterricht sowie in der Konfrontation mit inkompatiblen Konzepten (vgl. Kapitel 2.2.2). Zwar ist der hier eingesetzte Fragebogen in keiner Weise mit einem Unterricht zu vergleichen, dennoch werden die Schüler mit verschiedenen Erklärungsansätzen wiederholt konfrontiert, zusätzlich wurde das individuelle Vorwissen bei der Aufgabenkonstruktion nicht berücksichtigt. Die Bildung von synthetischen Modellen könnte also sogar durch das Antwortformat bedingt sein. Ob die synthetischen Modelle bereits a priori von den Schülern so konstruiert worden waren oder ad hoc von ihnen gebildet wurden, könnte nur durch weitere, ggf. qualitative Studien, zu Tage gebracht werden.

Insgesamt bleibt eine Interpretation der Befunde dieser Arbeit entweder zugunsten des Kohärenzansatzes oder des Fragmentierungsansatzes schwierig: Keine der beiden Perspektiven kann alle Befunde zufriedenstellend erklären. Hingegen können die Theorien differenziell je nach Themenbereich zur Interpretation der Ergebnisse genutzt werden. Damit kommt diese Arbeit zu einem ähnlichen Schluss wie bereits andere Studien zuvor, in denen die verschiedenen Ansätze differenziert nach Personensubgruppen zur Interpretation von Antwortmustern genutzt wurden (vgl. Kleickmann et al., 2011; Schneider & Hardy, 2013). Diese Nutzung multipler Theorien widerspricht allerdings dem Prinzip der sparsamen Erklärungen, da so insgesamt viele theoretische Annahmen zur Erklärung der empirischen Evidenzen notwendig sind. Erstrebenswert wäre also eine integrierte Theorie, die umfassend in der Lage ist, die verschiedenen Befunde sowohl hinsichtlich unterschiedlicher Themenbereiche, unterschiedlicher Personengruppen und sogar unterschiedlicher Antwortformate, zu interpretieren (vgl. diSessa, 2006) und die im Idealfall zuverlässige Vorhersagen über das zu erwartende Antwortverhalten zulässt.

Diese Forderung ist allerdings so weitreichend, dass sie aufgrund der notwendigen empirischen Arbeiten und der augenscheinlichen Unvereinbarkeit der Vertreter der beiden Ansätze, unmöglich erscheint. Studien, wie die vorliegende oder auch weiterführende, können allerdings einen Ertrag in diese Richtung leisten.

## **6.5 Ausblick**

Die Resultate meiner Arbeit geben vielfältigen Anlass zu weiteren Studien. Zunächst sind methodische Veränderungen und Erweiterungen der hier beschriebenen Studien denkbar: Um einen möglichen Einfluss der Aufgabenstämme unabhängig von den zu einem Stamm genesteten Items zu untersuchen, könnten die Aufgaben im Sinne eines Mehrebenenmodells zerlegt werden, d.h. für Items und Aufgabenstämme sollten separate Varianzkomponenten eingeführt werden. Dabei wäre es für die Untersuchung des Einflusses von Kontextmerkmalen notwendig, dass neben dem Einfluss der individuellen Aufgabenstämme feste Effekte auf diese untersucht werden können.

Neben einer Überlagerung des Kontexteffektes durch die einzelnen Items könnte eine differenzielle Wirkung von situativen Merkmalen bei individuellen Personen oder Personengruppen einen verschleiern Einfluss gehabt haben. Eine Analyse der Wirkung bei verschiedenen Subpopulationen oder sogar Einzelpersonen könnte darüber Aufschluss bringen.

Qualitative Untersuchungen zur Wahrnehmung der Aufgabenstämme könnten helfen aufzuklären, welche Merkmale der Kontexte überhaupt für die Probanden salient sind. Zwar wurden in dieser Arbeit bereits mehrere verschiedene Merkmale betrachtet, allerdings besteht die Möglichkeit, dass von Grundschulkindern andere Faktoren zur Beschreibung der naturwissenschaftlichen Phänomene genutzt werden. Denkbar wären hierbei z.B. Interviewstudien. Verschiedene Situationen, in denen z.B. ein Verdunstungsphänomen beschrieben wird, könnten von den Probanden nach Ähnlichkeit und Unähnlichkeit sortiert werden, während gleichzeitig nach den Gründen für die Sortierung gefragt wird.

In der vorliegenden Arbeit konnte kein Hinweis auf eine zunehmende Dekontextualisierung der Vorstellungen gefunden werden. Allerdings waren die hier beschriebenen Untersuchungen querschnittlich angelegt. Längsschnittliche Studien, bei denen die Veränderung des Wissensgefüges von individuellen Personen nachverfolgt werden kann, könnten solche Entwicklungsverläufe besser zu Tage bringen. Dafür sind sowohl qualitative wie auch quantitative Vorgehensweisen denkbar.

Desweiteren könnten die Ergebnisse dieser Arbeit, v.a. zu den Bereichen Dichte und Verdrängung, auf andere Aufgabenarten übertragen werden, um so die Reichweite der Befunde besser einschätzen zu können.

Da die Ergebnisse der beiden Studien dieser Arbeit für die beiden untersuchten Themenbereiche jeweils unterschiedlich ausfielen, wäre eine Übertragung der Vorgehensweise auf andere

Themenbereiche erstrebenswert. Solche Untersuchungen könnten langfristig Aufschluss darüber bringen, welche inhaltlich bedeutsamen Aufgabenmerkmale für die Grundschule schwierigkeiterzeugend sind. Dazu liegen für den Grundschulbereich bislang nur wenige empirische Untersuchungen vor.

Neben methodischen Abwandlungen der Arbeit und der Übertragung auf andere Themenbereiche scheint auch die Untersuchung des Zusammenhangs der hier gefundenen Ergebnisse mit anderen theoretischen Bereichen und deren Übertragung auf praktische Bereiche angezeigt. Zunächst beschäftigten sich die beschriebenen Untersuchungen mit der Ausgestaltung von Messverfahren zur Erfassung von Kompetenzen, sowie einer möglichen Übertragung der Befunde für die Beschreibung von Kompetenzen (siehe Kapitel 6.3). In einem nächsten Schritt könnten die Befunde, wie oben bereits angedeutet, sowohl für die Diagnostik wie auch die Förderung von Kompetenzen weitere Verwendung finden (vgl. Klieme & Leutner, 2006). Besonders das Aufgabenmerkmal Kongruenz kann zu einer vollständig Diagnostik von Schülerkompetenzen in den Bereichen Dichte und Verdrängung verhelfen. Darüber hinaus bietet die gezielte Variation der Schwierigkeit von Lernaufgaben durch dieses Merkmal bei der Förderung von Kompetenzen Möglichkeiten zur Differenzierung zwischen Schülern mit unterschiedlichem Wissensstand (vgl. Fischer & Draxler, 2002). Sogar eine sinnvolle Progression von Lernaufgaben könnte mit Hilfe dieses Merkmals gestaltet werden.

Das Aufgabenmerkmal der Kongruenz wird, neben der Dichte, bereits im Bereich Mathematik eingesetzt: Das Merkmal wurde eingesetzt in einem Test, der den *natural number bias*, also die Verzerrung kindlicher Konzeptionen von rationalen Zahlen durch das zuvor gelernte Konzept von natürlichen Zahlen, erfassen soll (Vosniadou & Vamvakoussi, 2006). Schwierigkeiten treten dabei u.a. im Vergleich der Größe von Bruchzahlen auf. Zur Erfassung dieser Schwierigkeit wurden kongruente Aufgaben, in denen die Anwendung des Konzeptes natürlicher Zahlen zu richtigen Antwort führte (z.B.  $0.745 > 0.36$ ) und inkongruente Aufgaben, in denen die Anwendung des Konzeptes zu fehlerhaften Antworten führte (z.B.  $0.125 > 0.87$ ), unterschieden und wiesen auch empirisch unterschiedliche Schwierigkeiten auf (Vamvakoussi, Van Dooren & Verschaffel, 2012; Van Hoof, 2012). Ob sich Lerneffekte in einem der beiden Bereiche (Dichte, rationale Zahlen) auch auf den jeweils anderen übertragen ist empirisch eine offene Frage.

Ein interessanter Zusammenhang der Konsistenz bzw. Inkonsistenz von Schülervorstellungen könnte mit epistemologischen Überzeugungen bestehen (z.B. Bromme, 2005; Perry, 1970; Schommer, 1990). In gängigen Ansätzen werden verschiedene Aspekte epistemologischer Überzeugungen dabei den Dimensionen *Natur des Wissens* und *Prozess des Wissenserwerbs* zugeordnet (z.B. Hofer & Pintrich, 1997). Denkbar wäre nun ein Zusammenhang zwischen Überzeugungen über die Natur des Wissens, in denen Wissen als feststehend im Gegensatz zu vorläufig feststehend oder veränderlich betrachtet wird, mit der Kontextualität von Vorstellungen bzw. Erklärungsansätzen. Besondere Parallelen scheinen darüber hinaus zwischen dem

Fragmentierungsansatz von diSessa (2006) und dem kontextuellen Ansatz epistemologischer Überzeugungen von Hammer und Ebly (2002) zu bestehen, bei denen die Überzeugungen als unabhängige kognitive Elemente (epistemologische Ressourcen) mit kontextspezifischem Aktivationsmuster beschrieben werden. Dabei kann sich, im Gegensatz zum hier untersuchten konzeptuellen Verständnis, eine Sophistizierung der epistemologischen Überzeugungen in einer flexiblen Differenzierung nach Domänen ausdrücken (vgl. Bromme, Kienhues & Stahl, 2008; Porsch, Bromme & Pollmeier, 2010). Mögliche Zusammenhänge zwischen der Kontextualisierung naturwissenschaftlicher Erklärungsansätze und der Kontextualisierung epistemologischer Ressourcen könnten untersucht werden.

## 7 Literatur

- AAAS (American Association for the Advancement of Science) (1993). *Benchmarks for Science Literacy*. Washington, DC: AAAS Project 2061.
- Andersen, E.B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123-140.
- Anderson, J.R., Reder, L.M., & Simon, H.A. (1996). Situated learning and education. *Educational Researcher*, 25, 5-11.
- Ausubel, D.P. (1968). *Educational psychology: a cognitive view*. New York: Holt, Rinehart and Winston.
- Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59 (4), 390-412.
- Bar, V. (1989). Children's views about the water cycle. *Science Education*, 73, 481-500.
- Bar, V., & Galili, I. (1994) Stages of children's views about evaporation. *International Journal of Science Education*, 16, 157-174.
- Bar, V., & Travis, A.S. (1991). Children's views concerning phase changes. *Journal of Research in Science Teaching*, 28, 363-382.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W. et al. (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske u. Budrich.
- Baumert, J., Bos, W., & Lehmann, R. (Hrsg.). (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. I. Mathematische und naturwissenschaftliche Bildung am Ende der Pflichtschulzeit*. Opladen: Leske u. Budrich.
- Beinbrech, C., & Möller, K. (2008). Entwicklung naturwissenschaftlicher Kompetenz im Sachunterricht. In H. Giest, A. Hartinger, & J. Kahlert (Hrsg.), *Forschungen zur Didaktik des Sachunterrichts. Bd. 7: Kompetenzniveaus im Sachunterricht* (S. 101-117). Bad Heilbrunn: Klinkhardt.
- Beveridge, M. (1985). The development of young children's understanding of the process of evaporation. *British Journal of Educational Psychology*, 55, 84-90.
- Biddulph, F., & Osborne, R. (1984). Pupils' ideas about floating and sinking. *Research in Science Education*, 14, 114-124.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31.
- Bonsen, M., Lintfort, K., Bos, W., & Frey, K.A. (2008). TIMSS 2007 Grundschule - Eine Einführung in die Studie. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter, & G. Walther, (Hrsg.) (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bögeholz, S., Höhle, C., Langlet, J., Sander, E., & Schlüter, K. (2004). Bewerten – Urteilen – Entscheiden im biologischen Kontext: Modelle in der Biologiedidaktik. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 89-115
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Aufl.). Berlin: Springer Verlag.
- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C., & Walther, G. (Hrsg.). (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster u.a.: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., Voss, A., & Walther, G. (Hrsg.). (2005). *IGLU. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G., & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bromme, R. (2005). Thinking and knowing about knowledge: A plea for and critical remarks on psychological research programs on epistemological beliefs. In M. Hoffmann, J. Lenhard, & F. Seeger, (Eds.), *Activity and sign - Grounding mathematics education* (pp. 191-201), New York: Springer.

- Bromme, R., Kienhues, D., & Stahl, E. (2008). Knowledge and epistemological beliefs: An intimate but complicate relationship. In M. S. Khine (Ed.), *Knowing, knowledge, and beliefs: Epistemological studies across diverse cultures* (pp. 423 - 444). New York: Springer.
- Bullock, M., Sodian, B., & Koerber, S. (2009). Doing experiments and understanding science: Development of scientific reasoning from childhood to adulthood. In W. Schneider & M. Bullock (Hrsg.), *Human development from early childhood to early adulthood: Findings from a 20 year longitudinal study* (pp. 173-197). New York, NY: Psychology Press.
- Butts, D.P., Hoffmann, H.M., & Anderson, M. (1993). Is hands-on experience enough? A study of young children's views of sinking and floating objects. *Journal of Elementary Science Education*, 5(1), 50-64.
- Bybee, R.W. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (Hrsg.), *Scientific Literacy* (S. 37-68). Kiel: IPN.
- Caravita, S., & Halldén, O. (1994). Re-framing the problem of conceptual change. *Learning and Instruction*, 4, 89-111.
- Caravita, S. (2001). A re-framed conceptual change theory? *Learning and Instruction*, 11, 421-429.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, Massachusetts: MIT Press.
- Carr, M., Kurtz, B.E., Schneider, W., Turner, L.A., & Borkowski, J.G. (1989). Strategy acquisition and transfer among American and German children: Environmental influences on metacognitive development. *Developmental Psychology*, 25, 765-771.
- Chi, M.T.H. (2008). Three types of conceptual change: belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Hrsg.), *International Handbook of Research on Conceptual Change* (pp. 61-82). New York: Routledge.
- Chi, M. T. H., Slotta, J. D., & DeLeeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, 4, 27-43.
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63, 1-49.
- Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen. *Unterrichtswissenschaft*, 31, 122-141.
- Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (1978). *Mind in society: The development of higher psychological processes*. L. S. Vygotsky. Oxford England: Harvard U Press.
- Coştu, B., & Ayas, A. (2005). Evaporation in different liquids: Secondary students' conceptions. *Research in Science & Technological Education*, 23, 75-97.
- De Boeck, P. (2008). Random item IRT models. *Psychometrika*, 73, 533-559.
- De Boeck, P. & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.
- De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F. et al. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*, 39, 1-28.
- diSessa, A.A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, 10, 105-225.
- diSessa, A.A. (2006). A history of conceptual change research: Threads and fault lines. In R.K. Sawyer & R.K. Sawyer (Hrsg.), *The Cambridge handbook of: The learning sciences* (pp. 265-281). New York, NY: Cambridge University Press.
- diSessa, A.A. (2008). A bird's-eye view of the "pieces" vs. "coherence" controversy (from the "pieces" side of the fence). In S. Vosniadou (Hrsg.), *International handbook of research on conceptual change* (pp. 35-60). New York: Routledge.
- diSessa, A.A., Gillespie, N.M., & Esterly, J.B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science: A Multidisciplinary Journal*, 28, 843-900.
- diSessa, A., & Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education*, 20, 1155-1191.
- Dole, J.A., & Sinatra, G.M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist*, 33, 109-128.
- Doran, H., Bates, D., Bliese, P., & Dowling, M. (2007). Estimating the multilevel Rasch model: With the lme4 package. *Journal of Statistical Software*, 20, 1-18.



- Draxler, D. (2006). *Aufgabendesign und basismodellorientierter Physikunterricht*. (unveröffentlichte Dissertation). Universität Duisburg-Essen, Essen.
- Driver, R., Squires, R., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science - Research into children's ideas*. New York.
- Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education*, 11, 481-490.
- Driver, R., & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education*, 5, 61-84.
- Duit, R. (1999). Conceptual change approaches in science education. In W. Schnotz, S. Vosniadou & M. Carretero (Hrsg.), *New perspectives on conceptual change* (pp. 263-282). New York: Pergamon Press.
- Embretson (Whitely), S.E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Engel Clough, E., & Driver, R. (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Science Education*, 70, 24.
- Engelen, A., Jonen, A., & Möller, K. (2002). Lernfortschrittsdiagnosen durch Interviews - Ergebnisse einer Pilotstudie zum "Schwimmen und Sinken" im Sachunterricht der Grundschule. In K. Spreckelsen, K. Möller, & A. Hartinger (Hrsg.), *Ansätze und Methoden empirischer Forschung zum Sachunterricht. Forschungen zur Didaktik des Sachunterrichts* (S. 155-173). Bad Heilbrunn: Klinkhardt.
- Fassouloupoulos, G., Kariotoglou, P., & Koumaras, P. (2003). Consistent and inconsistent pupils' reasoning about intensive quantities: The case of density and pressure. *Research in Science Education*, 33, 71-87.
- Fischer, G.H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.
- Fischer, H. (2008). "Und die leichten Sachen sind oben geblieben." Wie deuten Kinder Phänomene des Schwimmens und Sinkens? Eine qualitative Untersuchung. *Sache, Wort, Zahl*, 36(94), 55-60.
- Fischer, H.E., & Draxler, D. (2006). Konstruktion und Bewertung von Physikaufgaben. In E. Kircher, R. Girwitz, & P. Häußler (Hrsg.), *Physikdidaktik. Theorie und Praxis* (S. 639-655). Berlin: Springer.
- Garnett, P.J., Garnett, P.J., & Hackling, M.W. (1995). Students' Alternative Conceptions in Chemistry: A Review of Research and Implications for Teaching and Learning. *Studies in Science Education*, 25, 69-96.
- Gesellschaft für Didaktik des Sachunterrichts (2002). *Perspektivrahmen Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Grygier, P. (2008). *Wissenschaftsverständnis von Grundschulern im Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Halldén, O., Scheja, M., & Haglund, L. (2008). The contextuality of knowledge: An intentional approach to meaning making and conceptual change. In Vonsiadou, S. (Ed.), *International handbook of research on conceptual change*. (pp. 509-532). London: Routledge.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle. Merkmale und ihre Bedeutung – dargestellt anhand von Kompetenzen beim Experimentieren. *MNU*, 57, 196-203.
- Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B.K. Hofer & P.R. Pintrich (Eds.), *Personal epistemology: the psychology of beliefs about knowledge and knowing* (pp. 169-190). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Hardy, I., Kleickmann, T., Koerber, S., Mayer, D., Möller, K., Pollmeier, J., Schwippert, K., & Sodian, B. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter. Projekt Science-P. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms*. (S. 115-125). Weinheim; Basel: Beltz.
- Hardy, I., Jonen, A., Möller, K. & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of 'floating and sinking.'. *Journal of educational psychology*, 98 (2), 307-326.
- Hardy, I., & Stern, E. (2011). Visuelle Repräsentationen der Dichte: Auswirkungen auf die konzeptuelle Umstrukturierung bei Grundschulkindern. *Unterrichtswissenschaft*, 39, 35-48.
- Hartig, J., & Frey, A. (2012). Konstruktvalidierung und Skalenbeschreibung in der Kompetenzdiagnostik durch die Vorhersage von Aufgabenschwierigkeiten. *Psychologische Rundschau*, 63, 43-49.

- Havu-Nuutinen, S. (2005). Examining young children's conceptual change process in floating and sinking from a social constructivist perspective. *International Journal of Science Education*, 27, 259-279.
- Hofer, B.K., & Pintrich, P.R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67, 88-140.
- Hsin, C., & Wu, H. (2011). Using scaffolding strategies to promote young children's scientific understandings of floating and sinking. *Journal of Science Education and Technology*, 20, 656-11; 666.
- Ionnides, C., & Vosniadou, S. (2002). The changing meaning of force. *Cognitive Science Quarterly*, 2, 5-61.
- Janssen, R., Schepers, J., & Peres, D. (2004). Models with Item and Item Group Predictors. In P. De Boeck & M. Wilson (Hrsg.), *Explanatory Item Response Models* (pp. 198-212). New York: Springer.
- Johnson, P. (1998). Children's understanding of changes of state involving the gas state, part 2: Evaporation and condensation below boiling point. *International Journal of Science Education*, 20, 695-709.
- Joung, Y.J. (2009). Children's typically-perceived-situations of floating and sinking. *International Journal of Science Education*, 31, 101-127.
- 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.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*. Berlin: Logos-Verl.
- Kleickmann, T., Hardy, I., Möller, K., Pollmeier, J., Tröbst, S., & Beinbrech, C. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 263-281.
- Kleickmann, T., Pollmeier, J., Hardy, I., & Möller, K. (2011). Die Struktur naturwissenschaftlichen Wissens von Grundschulkindern – eine person- und variablenzentrierte Analyse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 200-212.
- Klieme, E. (2000). Fachleistungen im voruniversitären Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzstufen und Unterrichtsschwerpunkte. In J. Baumert, W. Bos, & R. Lehmann (Eds.), *Dritte internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (S. 57-128). Leverkusen: Leske + Budrich.
- Klieme, E., Artelt, C., & Hartig, J. (Hrsg.). (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Pädagogik*, 52, 876-903.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.E., & Vollmer, H.J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Frankfurt, Main: Deutsches Institut für Internationale Pädagogische Forschung.
- Labudde, P. (2007). How to develop, implement and assess standards in science education? 12 challenges from a Swiss perspective. In D. Waddington, P. Nentwig & S. Schanze (Hrsg.), *Making it comparable: Standards in science education* (pp. 277-301). Münster: Waxmann.
- Labudde, P., & Adamina, M. (2008). HarmoS Naturwissenschaften: Impulse für den naturwissenschaftlichen Unterricht von morgen. *Beiträge zur Lehrerbildung*, 26, 351-360.
- Lankes, E.M., Bos, W., Mohr, I., Plaßmeier, N., Schwippert, K., Sibberns, H., & Voss, A. (2003). Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU) und ihrer Erweiterung um Mathematik und Naturwissenschaften (IGLU-E). In W. Bos, E.M. Lankes, M. Prenzel, K. Schwippert, G. Walther, & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 7-27). Münster: Waxmann.
- Larsson, Å., & Halldén, O. (2010). A structural view on the emergence of a conception: Conceptual change as radical reconstruction of contexts. *Science Education*, 94, 640-664.
- Lenhard, W., & Schneider, W. (Hrsg.). (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58, 103-117.

- Maier, U. (2010). Formative Assessment - ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, 13, 293-308.
- McClelland, D.C. (1973). Testing for competence rather than for 'intelligence.'. *American Psychologist*, 28, 1-14.
- Ministerium für Schule und Weiterbildung NRW (Hrsg.) (2008). *Richtlinien und Lehrpläne für die Grundschulen in Nordrhein-Westfalen – Lehrplan Sachunterricht*. Frechen: Ritterbach Verlag.
- Möller, K. (1999). Konstruktivistisch orientierte Lehr-Lernprozessforschung im naturwissenschaftlich-technischen Bereich des Sachunterrichts. In W. Köhnlein (Hrsg.), *Vielperspektivisches Denken im Sachunterricht* (S. 125-191). Bad Heilbrunn: Klinkhardt.
- Möller, K. (2004). *Verstehen durch Handeln beim Lernen naturwissenschaftlicher und technikbezogener Sachverhalte*. Bad Heilbrunn: Klinkhardt.
- Möller, K. (2010). Lernen von Naturwissenschaften heisst: Konzepte verändern. In: Labudde, P. (Hrsg.): *Fachdidaktik Naturwissenschaft. 1.-9. Schuljahr* (S. 57-72). Bern: Haupt.
- Morrison, C.M. (2003). Interpret with caution: multicollinearity in multiple regression of cognitive data. *Perceptual and Motor Skills*, 97, 80-82.
- National Science Education Standards. (1996). *National Research Council, National Committee on Science Education Standards and Assessment*. Washington, DC.
- Novak, J.D. (1987). *Misconceptions and educational strategies in science and mathematics. Proceedings of the International Seminar Volume I*. Washington, D.C.: ERIC Clearinghouse.
- Osborne, R.J., & Cosgrove, M.M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching*, 20, 825-838.
- Perry, W.G.J. (1970). *Forms of intellectual and ethical development in the college years*. Oxford, England: Holt, Rinehart & Winston.
- Piaget, J. (2001). *The child's conception of physical causality* (new material ed.). New Brunswick, NJ: Transaction Publishers.
- Pintrich, P.R., Marx, R.W., & Boyle, R.A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63, 167-199.
- PISA-Konsortium Deutschland (Hrsg.) (2007). *PISA 06 - Die Ergebnisse der dritten Vergleichsstudie*. Münster: Waxmann
- PISA-Konsortium Deutschland (Hg.) (2008). *PISA 2006 in Deutschland - Die Kompetenzen der Jugendlichen im dritten Ländervergleich*. Münster: Waxmann
- Pollmeier, J. (2014). Die Erfassung des Dichteverständnisses von Grundschulern mit schriftlichen Testaufgaben. *Vortrag auf der 2. Tagung der Gesellschaft für empirische Bildungsforschung*. Frankfurt a.M.
- Pollmeier, J., Hardy, I., Koerber, S., & Möller, K. (2011). Lassen sich naturwissenschaftliche Lernstände im Grundschulalter mit schriftlichen Aufgaben valide erfassen? *Zeitschrift für Pädagogik*, 57, 834-853.
- Porsch, T., Bromme, R., & Pollmeier, J. (2010). Was muss man tun um sicher die richtige Lösung zu finden? - Quellenpräferenzen von Grundschulkindern in verschiedenen Fachkontexten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 42, 90-98.
- Posner, G.J., Strike, K.A., Hewson, P.W., & Gertzog, W.A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66, 211-227.
- Prenzel, M., Häußler, P., Rost, J., & Senkbeil, M. (2002). Der PISA-Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten vorhersagen? *Unterrichtswissenschaft*, 30, 120-135.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000 – Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 191-248). Opladen: Leske + Budrich.
- Prenzel, M., Geiser, H., Langeheine, R., & Lobemeier, K. (2003). Das naturwissenschaftliche Verständnis am Ende der Grundschule. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther, & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 143-187). Münster: Waxmann.

- Qualifications and Curriculum Authority. (2000). *Department for Education and Employment (DfEE). Science: The National Curriculum for England: Key Stages 1-4.*
- Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.* Oxford England: Nielsen & Lydiche.
- Rönebeck, S., Schöps, K., Prenzel, M., & Hammann, M. (2008). Naturwissenschaftliche Kompetenz im Ländervergleich. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2006 in Deutschland – die Kompetenzen der Jugendlichen im dritten Ländervergleich* (S. 67-94). Münster: Waxmann.
- Rosenshine, B. (1970). Evaluation of classroom instruction. *Review of Educational Research*, 40, 279-300.
- Rost, J. (1996). *Lehrbuch Testtheorie - Testkonstruktion.* Bern: Huber.
- Rost, J., Prenzel, M., Carstensen, C.H., Senkbeil, M., & Groß, K. (2004). *Naturwissenschaftliche Bildung in Deutschland. Methoden und Ergebnisse von PISA 2000.* Wiesbaden: VS Verl. f. Sozialwiss.
- Russell, T., Harlen, W., & Watt, D. (1989). Children's ideas about evaporation. *International Journal of Science Education*, 11, 566-576.
- Schecker, H., & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45-66.
- Schmidt, M., & Schecker, H. (2006). Kompetenzmodellierung im Themenbereich Energie – Entwicklung eines Testinventars. In A. Pitton (Ed.), *Lehr und Lernen mit neuen Medien* (Vol. 26, pp. 108-110). Berlin: Lit Verlag.
- Schneider, M., & Hardy, I. (2013). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology*, 49, 1639-1649.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82, 498-504.
- Schwartz, G. (1978). Estimating the Dimensions of a Model. *The Annals of Statistics*, 6, 461-464.
- Scott, P., Asoko, H., & Leach, J. (2007) Students conceptions and conceptual learning in science. In S.K. Abell & N.G. Lederman (Eds.) *Handbook of Research on Science Education* (pp. 31-56). New Jersey: Lawrence Erlbaum.
- Sekretariat der ständigen Kultusministerkonferenz. (2005). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss. Beschluss vom 16.12.2004.* München: Wolters Kluwer.
- Smith, C. (2007). Bootstrapping processes in the development of students' commonsense matter theories: Using analogical mappings, thought experiments, and learning zo measure to promote conceptual change. *Cognition and Instruction*, 25, 337-398.
- Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, 21, 177-237.
- Smith, C., Maclin, D., Grosslight, L., & Davis, H. (1997). Teaching for understanding: A study of students' preinstruction theories of matter and a comparison of the effectiveness of two approaches to teaching about matter and density. *Cognition and Instruction*, 15, 317-393.
- Stark, R. (2003). Conceptual Change: Kognitiv oder situiert? *Zeitschrift für Pädagogische Psychologie*, 17, 133-144.
- Straatemeier, M., van der Maas, H.L J., & Jansen, B.R J.(2008). Children's knowledge of the earth: A new methodological and statistical approach. *Journal of Experimental Child Psychology*, 100, 276-296.
- Treagust, D.F., & Duit, R. (2008). Conceptual change: A discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education*, 3, 297-328.
- Tytler, R. (2000). A comparison of year 1 and year 6 students' conceptions of evaporation and condensation: Dimensions of conceptual progression. *International Journal of Science Education*, 22, 447-467.
- Tytler, R. (1998). The nature of students' informal science conceptions. *International Journal of Science Education*, 20, 901-927.
- Tytler, R., & Peterson, S. (2004). From "try it and see" to strategic exploration: Characterizing young children's scientific reasoning. *Journal of Research in Science Teaching*, 41, 94-118.
- Tytler, R., Peterson, S., & Prain, V. (2006). Picturing evaporation: Learning science literacy through a particle representation. *Teaching Science: The Journal of the Australian Science Teachers Association*, 52, 12-17.
- Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2012). Naturally biased? In search for reaction time evidence for a natural number bias in adults. *The Journal of Mathematical Behavior*, 31, 344-355.

- Van Hoof, J. (2012). Inhibition of the natural number bias in tasks with rational numbers: towards a comprehensive instrument. *Presentation on the 8th International Conference on Conceptual Change of the European Association for Research on Learning and Instruction in Trier, Germany.*
- Viering, T., Fischer, H.E., & Neumann, K. (2010). Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I. Projekt Physikalische Kompetenz. In E. Klieme, D. Leutner, & M. Kenk (Hrsg.), *Kompetenzmodellierung. Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms.* (S. 92-103). Weinheim; Basel: Beltz.
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4, 45-69.
- Vosniadou, S. (2012). The problem of conceptual change in the learning of science and mathematics. *Presentation on the 8th International Conference on Conceptual Change of the European Association for Research on Learning and Instruction in Trier, Germany.*
- Vosniadou, S., & Brewer, W.F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535-585.
- Vosniadou, S., & Ioannides, C. (1998). From conceptual development to science education: A psychological point of view. *International Journal of Science Education*, 20, 1213-1230.
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 11, 381-419.
- Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 3-34). New York, NY: Routledge.
- Walpuski, M., Kauertz, A., Kampa, N., Fischer, H.E., Mayer, J., Sumfleth, E. et al. (2010). ESNaS – Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In A. Gehrmann, U. Hericks, & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle – Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 171– 184). Bad Heilbrunn: Julius Klinkhardt.
- Weinert, F.E. (2001). Concept of competence: a conceptual clarification.. In D.S. Rychen & L.H. Salganik (Hrsg.), *Defining and selecting key competencies.* (S. 45-66). Göttingen: Hogrefe.
- Weinert, F.E. (Hrsg.). (2002). *Leistungsmessungen in Schulen. 2. unveränd. Aufl.* Weinheim: Beltz.
- Weiß, R.H. (2006). *Grundintelligenztest Skala 2 - Revision (CFT 20-R)*. Göttingen: Hogrefe.
- Whitely, S.E. (1980) Multicomponent latent trait models for ability tests. *Psychometrika*, 45, 479-494.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.
- Wittwer, J., Saß, S., & Prenzel, M. (2008). Naturwissenschaftliche Kompetenz im internationalen Vergleich: Testkonzeption und Ergebnisse. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter, & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenz von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 87-124). Münster: Waxmann.
- Wu, M., Adams, R., & Wilson, M. (2005). *ACER ConQuest*: Australian Council for Educational Research.
- Zeyer, A., Adamina, M., Gingings, F., & Labudde, E. (2008). HarmoS Naturwissenschaften: Entwicklung Umsetzung und Assessment von Standards im naturwissenschaftlichen Unterricht der Schweiz. In D. Höttecke (Hrsg.), *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung* (S. 206-208). Berlin: LIT Verlag Dr. Hopf.

## 8 Anhang

### Verzeichnis der Anhänge

A01	Rotationsdesign zur Zusammenstellung der Testhefte in Studie 1 .....	136
A01.1	Rotationsdesign für Klasse 2 .....	136
A01.2	Rotationsdesign für Klasse 3 und 4 .....	136
A02	Übersicht über eingesetzte Aufgaben und Items in der Querschnittstudie nach der Selektion anhand der psychometrischen Kennwerte .....	137
A03	Kennwerte aller Items aus Studie 1 .....	138
A04	Kontingenztabellen zu den Abhängigkeiten der Kontextmerkmale und der Niveaus untereinander .....	142
A04.1	Abhängigkeiten der Aufgabenmerkmale im Themenbereich Verdunstung/Kondensation .....	142
A04.2	Abhängigkeiten der Aufgabenmerkmale im Themenbereich Schwimmen/Sinken .....	144
A05	Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für alle Items des Themenbereichs Verdunstung/Kondensation ...	145
A05.1	Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Items zu Verdunstung .....	146
A05.2	Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Items zu Kondensation .....	147
A06	Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Items zu Auftrieb .....	148
A07	Kennwerte aller Items aus Studie 2 .....	149
A07.1	Kennwerte aller Items zu Verdunstung aus Studie 2 .....	149
A07.2	Kennwerte aller Items zu Kondensation aus Studie 2 .....	153
A07.3	Kennwerte aller Items zu Dichte und Verdrängung aus Studie 2 .....	156
A08	Hierarchischer Vergleich der Modelle mit Haupteffekten der Kontextmerkmale und Interaktionen der Kontextmerkmale für Verdunstung .....	157
A09	Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Verdunstung .....	158
A10	Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Kondensation .....	159

## Anhang A01

Rotationsdesign zur Zusammenstellung der Testhefte in Studie 1

### Anhang A01.1

Rotationsdesign für Klasse 2

Block	1a	2a	3a	4a	5a	6a	7	1b	2b	3b	4b	5b	6b	Anzahl Aufgaben gesamt
Anzahl Aufgaben	2	2	2	2	2	2	10	6	6	6	6	6	6+1	59
Testheft 1	2						10	6						18
Testheft 2		2					10		6					18
Testheft 3			2				10			6				18
Testheft 4				2			10				6			18
Testheft 5					2		10					6		18
Testheft 6						2	10						6+1	19
Anzahl Aufgaben gesamt	2	2	2	2	2	2	60	6	6	6	6	6	7	

### Anhang A01.2

Rotationsdesign für Klasse 3 und 4

Block	1a	2a	3a	4a	5a	6a	7	1b	2b	3b	4b	5b	6b	Anzahl Aufgaben gesamt
Anzahl Aufgaben	2	2	2	2	2	2	10	6+1	6+1	6+1	6+1	6+2	6+2	66
Testheft 1	2						10	6+1	3					22
Testheft 2		2					10		6+1	3				22
Testheft 3			2				10			6+1	3			22
Testheft 4				2			10				6+1	3		22
Testheft 5					2		10					6+2	2	22
Testheft 6						2	10	2					6+2	22
Anzahl Aufgaben gesamt	2	2	2	2	2	2	60	9	9	9	9	9	9	

## Anhang A02

Übersicht über eingesetzte Aufgaben und Items in der Querschnittstudie nach der Selektion anhand der psychometrischen Kennwerte

Konzept	Aufgabenformat							Gesamt
	<i>forced choice</i>		<i>multiple select</i>	<i>multiple choice</i>		offen	andere	
	NV-	ZV-		3 <sup>b</sup>	4 <sup>b</sup>			
	ZV <sup>a</sup>	WV						
Verdunstung	5/ 5 <sup>c</sup>	1/ 1	7/ 26	3/ 3	1/ 1	-/ -	-/ -	17/ 36
Kondensation	5/ 5	4/ 4	7/ 20	1/ 1	-/ -	-/ -	-/ -	17/ 30
Auftrieb	3/ 3	2/ 2	11/ 33	-/ -	-/ -	1/ 2	-/ -	17/ 40
Dichte	1/ 1	-/ -	-/ -	8/ 8	-/ -	-/ -	-/ -	9/ 9
Verdrängung	3/ 3	-/ -	2/ 7	-/ -	-/ -	-/ -	1/ 3	6/ 13
Gesamt	17/17	7/ 7	26/ 86	12/12	1/ 1	1/ 2	1/ 3	66/128

*Anmerkungen.* NV, naive Vorstellung; ZV, Zwischenvorstellung; WV, wissenschaftliche Vorstellung;

<sup>a</sup>Hierunter sind Items subsumiert, die ein naive gegen eine wissenschaftliche Vorstellung testen. <sup>b</sup>Anzahl vorgegebenen Antwortalternativen; <sup>c</sup>Anzahl Aufgabenstämme/Anzahl Items;



**Anhang A03****Kennwerte aller Items aus Studie 1**

Item	Be- reich	Niveau	N	M(SD)	estimate <sup>a</sup>	error <sup>a</sup>	weighted MNSQ <sup>a</sup>	Trenn- schärfe <sup>a</sup>
b10	Vd	ZV	445	0.90(0.30)	-1.89	0.05	0.98	0.29
b11	K	WV	446	0.34(0.48)	1.22	0.05	1.03	0.33
b12	A	ZV	447	0.53(0.50)	0.45	0.05	1.30	0.12
b13_1	A	NV	609	0.39(0.49)	1.21	0.05	0.98	0.48
b13_2rec <sup>b</sup>	A	ZV	609	0.52(0.50)	0.58	0.05	1.12	0.33
b13_3	A	NV	609	0.51(0.50)				
b13_4	A	WV	606	0.53(0.50)	0.51	0.05	1.25	0.18
b14	V	WV	447	0.73(0.45)	-0.61	0.05	0.94	0.50
b15	D	WV	446	0.72(0.45)	-0.53	0.05	0.96	0.49
b16di	Vd	WV	606	0.55(0.50)	0.29	0.04	0.94	0.48
b17_1	K	NV	426	0.40(0.49)	0.97	0.05	0.99	0.36
b17_2	K	WV	427	0.78(0.41)	-0.90	0.05	1.03	0.27
b17_3	K	NV	425	0.82(0.38)	-1.18	0.05	1.01	0.30
b17_4	K	WV	424	0.49(0.50)	0.55	0.05	1.00	0.38
b18_1rec	K	ZV	293	0.39(0.49)	1.15	0.05	0.95	0.46
b18_2	K	NV	293	0.58(0.50)	0.30	0.05	1.04	0.32
b18_3	K	NV	293	0.82(0.38)	-1.04	0.05	1.02	0.27
b18_4	K	WV	293	0.78(0.42)	-0.74	0.05	0.97	0.36
b20	K	ZV	274	0.39(0.49)	0.86	0.05	1.09	0.20
b21	Vd	WV	274	0.89(0.32)	-1.83	0.05	1.04	0.20
b22_1	K	WV	565	0.57(0.50)	0.20	0.04	1.02	0.35
b22_2rec	K	ZV	565	0.34(0.48)				
b22_3	K	NV	564	0.24(0.43)	1.79	0.05	1.04	0.29
b23	Vd	WV	565	0.51(0.50)	0.49	0.04	1.02	0.34
b24	K	WV	274	0.40(0.49)	0.81	0.05	0.99	0.44
b25	D	WV	274	0.66(0.47)	-0.25	0.05	0.92	0.52
b26_1rec	A	ZV	272	0.66(0.48)	-0.24	0.05	1.15	0.27
b26_2	A	NV	273	0.89(0.32)	-1.83	0.05	0.96	0.32
b26_3	A	WV	271	0.63(0.48)	-0.11	0.05	1.15	0.27
b26_4	A	NV	272	0.73(0.45)	-0.63	0.05	1.08	0.31
b27	A	WV	274	0.40(0.49)	1.06	0.05	1.19	0.27
b28_1	V	NV	472	0.94(0.23)				
b28_2	V	NV	472	0.75(0.44)	-0.46	0.05	1.07	0.35
b28_3	V	NV	472	0.52(0.50)	0.75	0.05	0.78	0.69
b28_4	V	WV	471	0.86(0.34)	-1.33	0.05	0.98	0.37
b30	Vd	ZV	307	0.73(0.45)	-0.68	0.05	1.02	0.33
b31	K	ZV	307	0.54(0.50)	0.20	0.05	1.01	0.41
b32	A	WV	306	0.68(0.47)	-0.34	0.05	1.07	0.35
b33	D	WV	483	0.63(0.48)	0.04	0.05	0.92	0.55
b34	V	WV	484	0.58(0.50)	0.32	0.05	0.81	0.65
b35_1rec	A	ZV	484	0.38(0.49)	1.27	0.05	1.26	0.19
b35_2	A	WV	482	0.43(0.50)				
b35_3	A	NV	482	0.35(0.48)				

b35_4	A	NV	480	0.84(0.37)	-1.22	0.05	1.06	0.25
b36_1rec	K	ZV	305	0.71(0.46)	-0.59	0.05	1.04	0.31
b36_2	K	NV	304	0.95(0.22)				
b36_3	K	WV	305	0.91(0.29)				
b37_1	Vd	NV	304	0.38(0.49)	0.95	0.05	1.07	0.25
b37_2	Vd	WV	304	0.39(0.49)	0.90	0.05	0.98	0.42
b37_3rec	Vd	ZV	303	0.50(0.50)	0.38	0.05	1.07	0.28
b37_4	Vd	NV	305	0.79(0.41)	-1.05	0.05	0.93	0.45
b38_1	A	WV	194	0.59(0.50)	0.50	0.05	1.21	0.21
b38_2	A	NV	195	0.84(0.37)	-1.03	0.05	1.14	0.17
b38_3rec	A	ZV	194	0.76(0.43)				
b38_4	A	NV	195	0.66(0.47)	0.13	0.05	1.02	0.41
b40	K	ZV	258	0.40(0.49)	0.86	0.05	1.14	0.14
b41	Vd	WV	255	0.67(0.47)	-0.41	0.05	1.07	0.22
b42	K	WV	433	0.36(0.48)	1.08	0.05	1.04	0.32
b43_1rec	Vd	ZV	260	0.62(0.49)	-0.12	0.05	0.91	0.53
b43_2	Vd	WV	259	0.59(0.49)	-0.02	0.05	0.91	0.56
b43_3	Vd	NV	260	0.25(0.43)	1.63	0.05	1.01	0.28
b43_4rec	Vd	ZV	259	0.27(0.45)	1.49	0.05	1.09	0.17
b44	K	WV	260	0.60(0.49)	-0.07	0.05	1.03	0.32
b45	D	WV	430	0.67(0.47)	-0.09	0.05	0.89	0.55
b46_1	A	NV	433	0.74(0.44)	-0.53	0.05	1.17	0.18
b46_2	A	WV	434	0.72(0.45)				
b46_3	A	NV	431	0.75(0.43)	-0.56	0.05	1.20	0.10
b46_4rec	A	ZV	431	0.76(0.43)	-0.63	0.05	1.17	0.13
b47_1	A	NV	259	0.39(0.48)	1.13	0.05	1.15	0.24
b47_2	A	NV	260	0.73(0.44)	-0.56	0.05	1.15	0.13
b47_3	A	NV	260	0.44(0.50)	0.88	0.05	1.10	0.30
b47_4	A	WV	259	0.71(0.45)	-0.46	0.05	1.03	0.34
b48_1	Vd	NV	192	0.26(0.44)	1.67	0.05	1.00	0.27
b48_2	Vd	WV	192	0.45(0.50)	0.76	0.05	0.96	0.47
b48_3	Vd	NV	192	0.63(0.48)	-0.05	0.05	1.09	0.22
b48_4	Vd	NV	191	0.85(0.36)	-1.37	0.05	0.98	0.33
b50	Vd	WV	272	0.62(0.49)	-0.22	0.05	1.04	0.34
b51	K	ZV	274	0.53(0.50)	0.21	0.05	1.05	0.32
b52	A	ZV	465	0.61(0.49)	0.01	0.05	1.13	0.26
b53	A	WV	273	0.53(0.50)	0.25	0.05	1.10	0.33
b54	D	WV	273	0.85(0.36)	-1.60	0.05	1.02	0.29
b55_1rec	A	ZV	274	0.47(0.50)				
b55_2	A	NV	273	0.20(0.40)	1.97	0.05	1.16	0.17
b55_3	A	WV	274	0.66(0.48)	-0.38	0.05	1.17	0.21
b55_4	A	NV	273	0.48(0.50)				
b56_1	K	ZV	465	0.93(0.25)				
b56_2	K	WV	465	0.72(0.45)	-0.60	0.05	0.95	0.41
b56_3	K	NV	463	0.98(0.15)				
b56_4	K	NV	465	0.96(0.20)				
b57_1	Vd	NV	463	0.72(0.45)	-0.61	0.05	0.96	0.37

b57_2rec	Vd	ZV	463	0.53(0.50)	0.29	0.04	1.11	0.19
b57_3	Vd	WV	462	0.81(0.40)	-1.14	0.05	0.94	0.43
b57_4	Vd	NV	465	0.88(0.32)	-1.78	0.05	1.00	0.28
b58_1rec	A	ZV	178	0.77(0.42)	-0.80	0.05	1.18	0.11
b58_2	A	NV	178	0.24(0.43)				
b58_3	A	WV	178	0.55(0.50)	0.34	0.05	1.20	0.18
b59	D	WV	179	0.41(0.49)	1.00	0.05	0.94	0.50
b60	Vd	ZV	257	0.82(0.38)	-1.34	0.05	0.91	0.48
b61_1	K	NV	256	0.22(0.42)	1.73	0.05	1.07	0.19
b61_2	K	WV	256	0.27(0.45)	1.45	0.05	0.97	0.39
b61_3	K	NV	256	0.66(0.48)	-0.36	0.05	1.10	0.18
b61_4	K	WV	255	0.54(0.50)	0.17	0.05	1.17	0.11
b62	Vd	ZV	257	0.62(0.49)	-0.19	0.05	0.97	0.42
b63	K	WV	257	0.47(0.50)	0.46	0.05	1.07	0.28
b64_1	Vd	ZV	256	0.75(0.43)				
b64_2	Vd	WV	257	0.53(0.50)	0.24	0.05	0.90	0.57
b64_3	Vd	NV	257	0.55(0.50)	0.11	0.05	1.04	0.30
b64_4	Vd	NV	256	0.54(0.50)	0.19	0.05	1.09	0.25
b65_1	A	WV	257	0.72(0.45)	-0.61	0.05	0.90	0.50
b65_2	A	WV	256	0.70(0.46)	-0.46	0.05	0.89	0.55
b65_3	A	WV	256	0.73(0.45)	-0.64	0.05	0.95	0.46
b65_4	A	WV	256	0.70(0.46)	-0.46	0.05	1.01	0.42
b66	V	WV	435	0.84(0.37)	-1.38	0.05	1.05	0.31
b67	D	WV	435	0.51(0.50)	0.53	0.05	0.90	0.55
b68_1	A	NV	255	0.80(0.40)	-1.11	0.05	0.93	0.44
b68_2	A	NV	254	0.44(0.50)				
b68_3	A	WV	254	0.57(0.50)	0.22	0.05	1.21	0.23
b68_4rec	A	ZV	255	0.73(0.45)	-0.64	0.05	1.13	0.25
b69_1	Vd	WV	162	0.59(0.49)	0.08	0.05	0.84	0.63
b69_2	Vd	NV	163	0.42(0.50)	0.87	0.05	0.94	0.49
b69_3	Vd	WV	159	0.72(0.45)				
b69_4rec	Vd	ZV	161	0.71(0.45)	-0.52	0.05	0.88	0.54
b70_L1	A	NV	1733	0.35(0.48)	1.31	0.04	1.16	0.27
b70_L2rec	A	ZV	1733	0.42(0.49)	0.93	0.04	1.09	0.36
b70_L3	A	WV	1733	0.03(0.18)				
b71	D	WV	1816	0.60(0.49)	0.08	0.04	1.03	0.42
b72_1	V	WV	1810	0.31(0.46)	1.48	0.04	0.77	0.67
b72_2	V	WV	1810	0.39(0.49)	1.08	0.04	0.86	0.60
b72_3	V	WV	1810	0.29(0.45)	1.63	0.04	0.75	0.70
b73_1	K	WV	1816	0.70(0.46)	-0.55	0.04	0.99	0.41
b73_2	K	NV	1813	0.76(0.43)	-0.85	0.04	0.96	0.42
b73_3	K	NV	1811	0.79(0.41)	-1.05	0.04	1.02	0.31
b73_4rec	K	ZV	1813	0.77(0.42)	-0.93	0.04	1.04	0.27
b74	Vd	WV	1814	0.58(0.49)	0.03	0.04	0.99	0.39
b75di	K	WV	1814	0.45(0.50)	0.63	0.37	1.03	0.33
b76_1	Vd	WV	1809	0.58(0.49)	0.04	0.04	0.93	0.51
b76_2	Vd	NV	1814	0.72(0.45)	-0.66	0.04	1.08	0.21

b76_3	Vd	NV	1815	0.77(0.42)	-0.92	0.04	0.95	0.41
b76_4rec	Vd	ZV	1810	0.83(0.37)	-1.36	0.04	0.94	0.42
b77_1	V	WV	1813	0.59(0.49)	0.12	0.04	0.82	0.64
b77_2	V	WV	1816	0.59(0.49)	0.12	0.04	0.82	0.65
b77_3	V	WV	1810	0.43(0.50)	0.86	0.04	0.87	0.60
b77_4	V	WV	1815	0.57(0.50)	0.22	0.04	0.82	0.65
b78_1	A	NV	1814	0.95(0.28)				
b78_2	A	WV	1808	0.77(0.42)				
b78_3	A	NV	1811	0.62(0.49)	-0.05	0.04	1.25	0.16
b78_4	A	NV	1812	0.82(0.39)	-1.20	0.04	1.09	0.25
b78_5rec	A	ZV	1809	0.72(0.45)	-0.56	0.04	1.20	0.17
b79	D	WV	1810	0.54(0.50)	0.34	0.36	0.91	0.56

*Anmerkungen.* NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; A, Auftrieb; D, Dichte; K, Kondensation; V, Verdrängung; Vd, Verdunstung; <sup>a</sup>Die Werte entstammen der jeweils eindimensionalen Rasch-Skalierung über den Themenbereich Verdunstung/Kondensation bzw. Schwimmen/Sinken. <sup>b</sup>Items, für die keine Rasch-Werte angegeben sind, wurden aufgrund ungenügender psychometrischer Qualität aus den Analysen ausgeschlossen.

## Anhang A03

Kontingenztabellen zu den Abhängigkeiten der Kontextmerkmale und der Niveaus untereinander

### Anhang A03.1

Abhängigkeiten der Aufgabenmerkmale im Themenbereich Verdunstung/Kondensation

Kontingenztable für alle Items des Themenbereichs Verdunstung/Kondensation

		Temperatur			Ort	
		kalt	Sonne	heiß	drinnen	draußen
Ort	drinnen	28	0	20		
	draußen	8	8	0		
Niveau	NV	13	4	5	14	8
	ZV	10	1	6	16	1
	WV	13	3	9	19	8

Kontingenztable für Items zu Verdunstung

		Temperatur			Ort		Wasserspiegel	
		kalt	Sonne	heiß	drinnen	draußen	ja	nein
Ort	drinnen	16	0	7				
	draußen	4	8	0				
Wasser- spiegel	ja	4	0	6	11	0		
	nein	16	8	1	13	12		
Niveau	NV	8	5	7	6	7	2	11
	ZV	5	1	3	9	1	4	6
	WV	7	4	2	9	4	5	8

Kontingenztable für Items zu Kondensation

		Temperatur			Ort		Kondensations- kern		Quelle	
		kalt	Sonne	heiß	drinnen	draußen	Fläche	Dampf	ja	nein
Ort	drinnen	12	-	13						
	draußen	4	-	0						
Konden- sations- kern	Fläche	14	-	12	24	2				
	Dampf	2	-	1	1	2				
Quelle	ja	9	-	12	21	4	20	1		
	nein	7	-	1	0	4	6	2		
Niveau	NV	5	-	4	8	1	9	0	8	1

<i>ZV</i>	5	-	2	7	0	7	0	4	3
<i>WV</i>	6	-	7	10	4	10	3	9	4

**Anhang A03.2**

Abhängigkeiten der Aufgabenmerkmale im Themenbereich Schwimmen/Sinken

Kontingenztafel für Items zu Auftrieb

		Größe		Gewicht		Hohl	
		klein	groß	leicht	schwer	hohl	gefüllt
Gewicht	leicht	29	1				
	schwer	0	5				
Hohl	hohl	18	2	19	1		
	gefüllt	11	4	11	9		
Niveau	NV	12	1	12	4	8	8
	ZV	8	2	8	2	4	6
	WV	9	3	10	4	8	6

## Anhang A05

Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für den Themenbereich Verdunstung/Kondensation

### Anhang A05.1

Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für alle Items des Themenbereichs Verdunstung/Kondensation

Untersuchtes Kontext- merkmal		Modell nur Haupteffekte		Modell Haupteffekte und Interaktion	
		B	SE	B	SE
Ort	(Intercept)	0.56*	0.22	0.59*	0.25
	ZV	0.00	0.30	-0.03	0.33
	WV	-0.16	0.26	-0.22	0.32
	Ort=draußen	-0.40	0.26	-0.48	0.40
	ZV*Ort=draußen			-0.04	1.01
	WV*Ort=draußen			0.17	0.55
	BIC	41825		41846	
$\chi^2$		$\chi^2(2)=0.12, p=.94$			
Temperatur	(Intercept)	0.54*	0.21	0.65**	0.25
	ZV	0.12	0.28	-0.05	0.37
	WV	-0.22	0.26	-0.40	0.34
	Temperatur=warm	-0.30	0.22	-0.57	0.38
	ZV*Temperatur=warm			0.41	0.57
	WV*Temperatur=warm			0.41	0.51
	BIC	41116		41136	
$\chi^2$		$\chi^2(2)=0.75, p=.69$			

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen;

<sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001;



## Anhang A05.2

Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Items zu Verdunstung

Untersuchtes Kontext- merkmal	Modell nur Haupteffekte		Modell Haupteffekte und Interaktion	
	B	SE	B	SE
Wasserspiegel(Intercept)	0.18	0.40	-0.92	0.60
Wasserspiegel=nein	0.36	0.35	1.66*	0.65
ZV	0.42	0.41	1.29 <sup>+</sup>	0.74
WV	0.07	0.38	1.79*	0.71
Wasserspiegel=nein*ZV			-0.92	0.85
Wasserspiegel=nein*WV				
V			-2.31**	0.81
BIC	21015		21027	
$\chi^2$	$\chi^2(2)=7.91, p=.02$			
Ort				
(Intercept)	0.78*	0.33	0.67 <sup>+</sup>	0.38
Ort=draußen	-0.54	0.35	-0.35	0.51
ZV	0.10	0.42	0.23	0.49
WV	-0.14	0.37	0.00	0.49
Ort=draußen*ZV			-0.47	1.10
Ort=draußen*WV			-0.33	0.76
BIC	21014		21033	
$\chi^2$	$\chi^2(2)=0.28, p=.87$			
Temperatur				
(Intercept)	0.66*	0.28	0.76*	0.30
Temperatur=Sonne	-0.41	0.37	-0.39	0.51
Temperatur=kalt	-0.58	0.40	-2.02*	0.88
ZV	0.43	0.39	0.38	0.47
WV	-0.15	0.35	-0.42	0.43
Temperatur=Sonne*ZV			-0.68	1.03
Temperatur=kalt*ZV			1.47	1.04
Temperatur=Sonne*WV			0.19	0.76
Temperatur=kalt*WV			2.09 <sup>+</sup>	1.10
BIC	20827		20862	
$\chi^2$	$\chi^2(4)=4.16, p=.38$			

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen;

<sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001;

### Anhang A05.3

Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Items zu Kondensation

Untersuchtes Kontext-merkmal		Modell nur Haupteffekte		Modell Haupteffekte und Interaktion	
		B	SE	B	SE
Temperatur	(Intercept)	0.42	0.32	0.54	0.38
	Temperatur=kalt	-0.11	0.33	-0.38	0.57
	ZV	-0.26	0.43	-0.48	0.53
	WV	-0.28	0.37	-0.41	0.51
	Temperatur=kalt*ZV			0.63	0.91
	Temperatur=kalt*WV			0.29	0.74
	BIC	20345		20364	
$\chi^2$		$\chi^2(2)=0.48, p=.79$			
Quelle	(Intercept)	-0.17	0.42	-1.43 <sup>+</sup>	0.75
	Quelle=nicht vorhanden	0.61 <sup>+</sup>	0.35	2.03*	0.80
	ZV	-0.05	0.43	1.61 <sup>+</sup>	0.87
	WV	-0.17	0.36	1.12	0.84
	Quelle=nicht vorhanden*ZV			-2.11*	0.99
	Quelle=nicht vorhanden*WV			-1.46	0.92
	BIC	20343		20358	
$\chi^2$		$\chi^2(2)=0.48, p=.12$			

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen;

<sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001;

## Anhang A06

Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Items zu Auftrieb

Untersuchtes Kontext-merkmal		Modell nur Haupteffekte		Modell Haupteffekte und Interaktion	
		B	SE	B	SE
Größe	(Intercept)	0.71***	0.20	0.81***	0.19
	Größe=groß	-0.21	0.33	-1.51*	0.67
	ZV	-0.19	0.30	-0.24	0.30
	WV	-0.11	0.29	-0.39	0.29
	Größe=groß*ZV			1.03	0.85
	Größe=groß*WV			2.03*	0.80
	BIC	23141		23155	
$\chi^2$		$\chi^2(2)=6.20, p=.05$			
Gewicht	(Intercept)	0.70***	0.20	0.81***	0.21
	Gewicht=schwer	-0.32	0.27	-0.79 <sup>+</sup>	0.41
	ZV	-0.15	0.30	-0.24	0.32
	WV	-0.05	0.27	-0.33	0.30
	Gewicht=schwer*ZV			0.32	0.69
	Gewicht=schwer*WV			1.04 <sup>+</sup>	0.59
	BIC	24488		24505	
$\chi^2$		$\chi^2(2)=3.07, p=.22$			
Hohl	(Intercept)	0.69**	0.21	0.77***	0.23
	Hohl=ja	-0.20	0.25	-0.42	0.38
	ZV	-0.15	0.30	-0.39	0.36
	WV	-0.07	0.27	-0.12	0.34
	Hohl=ja*ZV			0.75	0.63
	Hohl=ja*WV			0.13	0.56
	BIC	24488		24507	
$\chi^2$		$\chi^2(2)=1.49, p=.48$			

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen;

<sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001;

**Anhang A07**

Kennwerte aller Items aus Studie 2

**Anhang A07.1**

Kennwerte der Items zu Verdunstung aus Studie 2

Item	Be- reich	Niveau	N	M(SD)	estimate <sup>a</sup>	error <sup>a</sup>	weighted MNSQ <sup>a</sup>	Trenn- schärfe <sup>a</sup>
b1_1	Vd	WV	158	0.42(0.49)	0.08	0.12	0.87	0.52
b1_2	Vd	ZV	158	0.27(0.45)	0.84	0.13	0.96	0.42
b1_3	Vd	ZV	158	0.57(0.50)	-0.62	0.12	0.90	0.48
b1_4	Vd	NV	158	0.65(0.48)	-1.02	0.13	1.17	0.16
b1_5 <sup>b</sup>	Vd	NV	157	0.55(0.50)				
b1_6	Vd	NV	158	0.53(0.50)	-0.42	0.12	1.10	0.24
b10_1	Vd	WV	158	0.39(0.49)	0.26	0.12	0.88	0.51
b10_2	Vd	ZV	157	0.21(0.41)	1.26	0.13	1.06	0.30
b10_3	Vd	ZV	158	0.58(0.49)	-0.67	0.12	1.03	0.33
b10_4	Vd	NV	158	0.50(0.50)	-0.28	0.12	1.09	0.28
b10_5	Vd	NV	158	0.5(0.50)				
b10_6	Vd	NV	158	0.45(0.50)	-0.05	0.12	0.98	0.41
b11_1	Vd	WV	158	0.52(0.50)	-0.37	0.12	0.93	0.45
b11_2	Vd	ZV	158	0.25(0.43)	1.01	0.13	1.10	0.23
b11_3	Vd	ZV	158	0.49(0.50)	-0.26	0.12	0.88	0.50
b11_4	Vd	NV	157	0.56(0.50)	-0.57	0.12	0.98	0.40
b11_5	Vd	NV	157	0.27(0.45)				
b11_6	Vd	NV	158	0.67(0.47)	-1.10	0.13	1.08	0.23
b12_1	Vd	WV	158	0.37(0.48)	0.35	0.13	0.97	0.40
b12_2	Vd	ZV	158	0.29(0.46)	0.75	0.13	1.05	0.30
b12_3	Vd	ZV	158	0.41(0.49)	0.13	0.12	0.96	0.43
b12_4	Vd	NV	157	0.62(0.49)	-0.83	0.12	1.15	0.18
b12_5	Vd	NV	157	0.29(0.46)				
b12_6	Vd	NV	158	0.25(0.44)	0.97	0.13	1.10	0.21
b13_1	Vd	WV	158	0.35(0.48)	0.44	0.13	1.00	0.36
b13_2	Vd	ZV	157	0.36(0.48)	0.40	0.13	1.03	0.34
b13_3	Vd	ZV	157	0.73(0.45)	-1.40	0.13	0.86	0.47
b13_4	Vd	NV	158	0.66(0.48)	-1.04	0.13	1.11	0.21
b13_5	Vd	NV	158	0.54(0.50)				
b13_6	Vd	NV	158	0.70(0.46)				
b14_1	Vd	WV	157	0.43(0.50)	0.03	0.12	0.87	0.52
b14_2	Vd	ZV	157	0.39(0.49)	0.21	0.12	1.15	0.21
b14_3	Vd	ZV	158	0.53(0.50)	-0.44	0.12	0.98	0.40
b14_4	Vd	NV	158	0.63(0.49)	-0.89	0.12	1.11	0.24
b14_5	Vd	NV	158	0.61(0.49)				
b14_6	Vd	NV	158	0.72(0.45)				
b15_1	Vd	WV	158	0.34(0.47)	0.50	0.13	0.88	0.52
b15_2	Vd	ZV	158	0.22(0.41)	1.21	0.13	1.03	0.32
b15_3	Vd	ZV	156	0.39(0.49)	0.23	0.12	0.93	0.47
b15_4	Vd	NV	158	0.56(0.50)	-0.59	0.12	1.13	0.22
b15_5	Vd	NV	158	0.70(0.46)				

b15_6	Vd	NV	158	0.22(0.41)	1.20	0.13	1.13	0.16
b16_1	Vd	WV	158	0.53(0.50)	-0.44	0.12	0.86	0.54
b16_2	Vd	ZV	158	0.28(0.45)	0.77	0.13	1.07	0.27
b16_3	Vd	ZV	158	0.72(0.45)	-1.35	0.13	0.98	0.34
b16_4	Vd	NV	158	0.51(0.50)	-0.33	0.12	0.93	0.46
b16_5	Vd	NV	158	0.70(0.46)				
b16_6	Vd	NV	158	0.66(0.48)				
b17_1	Vd	WV	158	0.35(0.48)	0.43	0.13	0.87	0.54
b17_2	Vd	ZV	158	0.32(0.47)	0.56	0.13	1.05	0.31
b17_3	Vd	ZV	158	0.47(0.50)	-0.15	0.12	0.82	0.58
b17_4	Vd	NV	158	0.61(0.49)	-0.83	0.12	1.05	0.30
b17_5	Vd	NV	158	0.23(0.42)	1.11	0.13	1.10	0.22
b17_6	Vd	NV	158	0.59(0.49)	-0.71	0.12	1.11	0.22
b18_1	Vd	WV	157	0.24(0.43)	1.01	0.13	0.85	0.54
b18_2	Vd	ZV	158	0.34(0.47)	0.49	0.13	1.06	0.30
b18_3	Vd	ZV	158	0.43(0.50)	0.02	0.12	0.93	0.45
b18_4	Vd	NV	158	0.59(0.49)	-0.74	0.12	1.13	0.19
b18_5	Vd	NV	158	0.27(0.44)				
b18_6	Vd	NV	158	0.37(0.48)	0.33	0.12	1.05	0.31
b19_1	Vd	WV	158	0.24(0.43)	1.03	0.13	0.85	0.52
b19_2	Vd	ZV	157	0.30(0.46)	0.69	0.13	1.07	0.28
b19_3	Vd	ZV	158	0.64(0.48)	-0.96	0.12	0.94	0.41
b19_4	Vd	NV	157	0.53(0.50)	-0.43	0.12	1.03	0.35
b19_5	Vd	NV	158	0.58(0.49)				
b19_6	Vd	NV	157	0.31(0.46)	0.62	0.13	0.96	0.40
b2_1	Vd	WV	157	0.36(0.48)	0.36	0.13	0.88	0.51
b2_2	Vd	ZV	157	0.33(0.47)	0.53	0.13	1.04	0.33
b2_3	Vd	ZV	156	0.59(0.49)	-0.71	0.12	0.90	0.46
b2_4	Vd	NV	157	0.60(0.49)	-0.75	0.12	1.18	0.15
b2_5	Vd	NV	157	0.50(0.50)				
b2_6	Vd	NV	157	0.68(0.47)				
b20_1	Vd	WV	157	0.43(0.50)	0.04	0.12	0.88	0.53
b20_2	Vd	ZV	157	0.32(0.47)	0.58	0.13	1.05	0.32
b20_3	Vd	ZV	157	0.54(0.50)	-0.46	0.12	0.89	0.50
b20_4	Vd	NV	158	0.69(0.46)	-1.22	0.13	1.13	0.19
b20_5	Vd	NV	158	0.48(0.50)				
b20_6	Vd	NV	155	0.30(0.46)	0.64	0.13	1.07	0.27
b21_1	Vd	WV	158	0.50(0.50)	-0.30	0.12	0.82	0.57
b21_2	Vd	ZV	158	0.23(0.42)	1.11	0.13	1.02	0.34
b21_3	Vd	ZV	158	0.56(0.50)	-0.57	0.12	0.97	0.40
b21_4	Vd	NV	158	0.46(0.50)	-0.10	0.12	1.04	0.32
b21_5	Vd	NV	157	0.56(0.50)				
b21_6	Vd	NV	158	0.68(0.47)	-1.15	0.13	1.07	0.29
b22_1	Vd	WV	156	0.35(0.48)	0.43	0.13	0.95	0.42
b22_2	Vd	ZV	158	0.29(0.46)	0.73	0.13	0.95	0.42
b22_3	Vd	ZV	158	0.63(0.48)	-0.93	0.12	0.87	0.49
b22_4	Vd	NV	158	0.48(0.50)	-0.21	0.12	1.07	0.29

b22_5	Vd	NV	158	0.63(0.49)				
b22_6	Vd	NV	158	0.43(0.50)	0.02	0.12	1.06	0.31
b23_1	Vd	WV	156	0.32(0.47)	0.58	0.13	0.88	0.51
b23_2	Vd	ZV	155	0.27(0.45)	0.85	0.13	1.02	0.35
b23_3	Vd	ZV	156	0.46(0.50)	-0.11	0.12	0.79	0.62
b23_4	Vd	NV	156	0.54(0.50)	-0.50	0.12	1.11	0.23
b23_5	Vd	NV	156	0.27(0.44)	0.86	0.13	1.11	0.20
b23_6	Vd	NV	156	0.50(0.50)	-0.29	0.12	1.12	0.22
b24_1	Vd	WV	157	0.47(0.50)	-0.16	0.12	0.90	0.49
b24_2	Vd	ZV	158	0.35(0.48)	0.39	0.13	1.02	0.36
b24_3	Vd	ZV	158	0.52(0.50)	-0.39	0.12	0.95	0.43
b24_4	Vd	NV	158	0.56(0.50)	-0.57	0.12	1.04	0.32
b24_5	Vd	NV	157	0.28(0.45)	0.79	0.13	1.16	0.13
b24_6	Vd	NV	156	0.43(0.50)	0.03	1.36	1.12	0.24
b3_1	Vd	WV	157	0.49(0.50)	-0.25	0.12	0.91	0.48
b3_2	Vd	ZV	158	0.44(0.50)	0.00	0.12	1.14	0.21
b3_3	Vd	ZV	158	0.53(0.50)	-0.44	0.12	0.88	0.52
b3_4	Vd	NV	156	0.44(0.50)	-0.03	0.12	1.16	0.20
b3_5	Vd	NV	158	0.65(0.48)				
b3_6	Vd	NV	155	0.61(0.49)	-0.81	0.12	1.11	0.22
b4_1	Vd	WV	157	0.47(0.50)	-0.16	0.12	0.87	0.52
b4_2	Vd	ZV	158	0.28(0.45)	0.78	0.13	1.08	0.27
b4_3	Vd	ZV	158	0.75(0.44)	-1.52	0.13	0.95	0.34
b4_4	Vd	NV	158	0.42(0.49)	0.10	0.12	1.06	0.31
b4_5	Vd	NV	158	0.59(0.49)				
b4_6	Vd	NV	158	0.66(0.47)	-1.07	0.13	1.11	0.22
b5_1	Vd	WV	158	0.54(0.50)	-0.46	0.12	0.85	0.53
b5_2	Vd	ZV	158	0.34(0.47)	0.51	0.13	1.14	0.22
b5_3	Vd	ZV	158	0.45(0.50)	-0.05	0.12	0.88	0.50
b5_4	Vd	NV	158	0.54(0.50)	-0.49	0.12	1.05	0.32
b5_5	Vd	NV	158	0.33(0.47)	0.54	0.13	1.19	0.13
b5_6	Vd	NV	158	0.75(0.44)				
b6_1	Vd	WV	157	0.39(0.49)	0.23	0.12	0.88	0.53
b6_2	Vd	ZV	158	0.31(0.46)	0.64	0.13	1.04	0.32
b6_3	Vd	ZV	158	0.46(0.50)	-0.11	0.12	0.94	0.44
b6_4	Vd	NV	158	0.60(0.49)	-0.76	0.12	0.99	0.37
b6_5	Vd	NV	158	0.24(0.43)	1.05	0.13	1.15	0.14
b6_6	Vd	NV	158	0.76(0.43)				
b7_1	Vd	WV	158	0.32(0.47)	0.58	0.13	0.92	0.45
b7_2	Vd	ZV	158	0.28(0.45)	0.82	0.13	0.99	0.37
b7_3	Vd	ZV	158	0.49(0.50)	-0.25	0.12	0.88	0.51
b7_4	Vd	NV	158	0.61(0.49)	-0.79	0.12	1.07	0.29
b7_5	Vd	NV	158	0.69(0.46)				
b7_6	Vd	NV	157	0.24(0.43)	1.09	0.13	0.97	0.38
b8_1	Vd	WV	158	0.37(0.49)	0.32	0.12	0.96	0.41
b8_2	Vd	ZV	158	0.29(0.46)	0.75	0.13	1.05	0.31
b8_3	Vd	ZV	158	0.45(0.50)	-0.05	0.12	0.86	0.54

b8_4	Vd	NV	158	0.72(0.45)	-1.37	0.13	1.04	0.27
b8_5	Vd	NV	158	0.56(0.50)				
b8_6	Vd	NV	157	0.24(0.43)	1.09	0.13	1.01	0.32
b9_1	Vd	WV	158	0.42(0.49)	0.10	0.12	0.90	0.50
b9_2	Vd	ZV	158	0.35(0.48)	0.45	0.13	1.03	0.36
b9_3	Vd	ZV	157	0.50(0.50)	-0.27	0.12	0.94	0.44
b9_4	Vd	NV	157	0.43(0.50)	0.02	0.12	1.03	0.33
b9_5	Vd	NV	158	0.57(0.50)				
b9_6	Vd	NV	157	0.48(0.50)	-0.18	0.12	1.11	0.25

*Anmerkungen.* NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; A, Auftrieb; D, Dichte; K, Kondensation; V, Verdrängung; Vd, Verdunstung; <sup>a</sup>Die Werte entstammen der eindimensionalen Rasch-Skalierung. <sup>b</sup>Items, für die keine Rasch-Werte angegeben sind, wurden aufgrund ungenügender psychometrischer Qualität aus den Analysen ausgeschlossen.

**Anhang A07.2****Kennwerte der Items zu Kondensation aus Studie 2**

Item	Be- reich	Niveau	N	M(SD)	estimate <sup>a</sup>	error <sup>a</sup>	weighted MNSQ <sup>a</sup>	Trenn- schärfe <sup>a</sup>
k01_ad <sup>b</sup>	K	NV	75	0.72(0.45)				
k01_fu	K	NV	77	0.84(0.37)				
k01_fz	K	NV	76	0.45(0.50)				
k01_l3	K	WV	77	0.51(0.50)	-0.36	0.18	1.20	0.28
k01_loc	K	ZV	77	0.84(0.37)	-2.48	0.21	1.17	0.14
k01_mat	K	ZV	77	0.40(0.49)	0.16	0.19	1.12	0.34
k02_fo	K	NV	77	0.66(0.48)				
k02_fu	K	NV	77	0.53(0.50)				
k02_fz	K	NV	77	0.22(0.42)				
k02_l3	K	WV	77	0.57(0.50)	-0.70	0.19	1.08	0.39
k02_loc	K	ZV	77	0.18(0.39)	1.48	0.20	0.91	0.39
k02_mat	K	ZV	77	0.34(0.48)	0.47	0.19	1.15	0.28
k03_ad	K	NV	77	0.78(0.42)				
k03_fu	K	NV	76	0.00(0.00)				
k03_fz	K	NV	77	0.29(0.45)	0.79	0.19	1.20	0.21
k03_l3	K	WV	77	0.43(0.50)	0.03	0.18	1.12	0.35
k03_loc	K	ZV	77	0.40(0.49)				
k03_mat	K	ZV	77	0.47(0.50)	-0.17	0.18	1.06	0.38
k04_fo	K	NV	77	0.34(0.48)				
k04_fu	K	NV	77	0.70(0.46)				
k04_fz	K	NV	76	0.61(0.49)				
k04_l3	K	WV	77	0.56(0.50)				
k04_loc	K	ZV	77	0.45(0.50)				
k04_mat	K	ZV	77	0.45(0.50)	-0.07	0.18	1.09	0.36
k05_fo	K	NV	77	0.52(0.50)				
k05_fu	K	NV	77	0.62(0.49)				
k05_fz	K	NV	76	0.41(0.49)				
k05_l3	K	WV	77	0.48(0.50)	-0.24	0.18	1.04	0.45
k05_loc	K	ZV	77	0.35(0.48)	0.43	0.19	1.23	0.20
k05_mat	K	ZV	77	0.22(0.42)	1.20	0.20	0.91	0.48
k06_ad	K	NV	77	0.69(0.47)				
k06_fu	K	NV	76	0.36(0.48)				
k06_fz	K	NV	76	0.36(0.48)				
k06_l3	K	WV	77	0.55(0.50)	-0.57	0.18	1.16	0.34
k06_loc	K	ZV	77	0.30(0.46)				
k06_mat	K	ZV	77	0.49(0.50)	-0.30	0.18	0.94	0.52
k07_ad	K	NV	76	0.59(0.49)				
k07_fu	K	NV	77	0.73(0.45)				
k07_fz	K	NV	76	0.30(0.46)				
k07_l3	K	WV	77	0.30(0.46)	0.71	0.19	0.97	0.46
k07_loc	K	ZV	77	0.40(0.49)	0.16	0.18	1.23	0.24
k07_mat	K	ZV	77	0.43(0.50)	0.01	0.19	1.16	0.30
k08_ad	K	NV	77	0.82(0.39)				



k08_fu	K	NV	77	0.66(0.48)				
k08_fz	K	NV	76	0.17(0.38)				
k08_l3	K	WV	77	0.64(0.48)	-1.06	0.19	1.11	0.40
k08_loc	K	ZV	77	0.71(0.45)	-1.51	0.19	1.30	0.17
k08_mat	K	ZV	77	0.25(0.44)	0.98	0.19	0.87	0.52
k09_fo	K	NV	77	0.31(0.47)				
k09_fu	K	NV	76	0.36(0.48)				
k09_fz	K	NV	77	0.13(0.34)	1.91	0.21	1.11	0.16
k09_l3	K	WV	77	0.45(0.50)	-0.11	0.18	0.86	0.59
k09_loc	K	ZV	77	0.23(0.43)	1.10	0.20	0.89	0.51
k09_mat	K	ZV	77	0.42(0.50)	0.06	0.19	0.96	0.49
k10_fo	K	NV	77	0.40(0.49)				
k10_fu	K	NV	77	0.58(0.50)				
k10_l3	K	WV	77	0.39(0.49)	0.22	0.19	0.94	0.51
k10_loc	K	ZV	77	0.49(0.50)	-0.31	0.18	1.08	0.40
k10_mat	K	ZV	77	0.42(0.50)	0.09	0.18	0.89	0.55
k11_ad	K	NV	77	0.66(0.48)				
k11_fu	K	NV	77	0.86(0.35)				
k11_fz	K	NV	77	0.19(0.40)				
k11_l3	K	WV	77	0.56(0.50)	-0.64	0.19	1.08	0.43
k11_loc	K	ZV	77	0.59(0.49)	-0.82	0.19	1.01	0.47
k11_mat	K	ZV	77	0.42(0.50)	0.09	0.18	0.76	0.67
k12_ad	K	NV	77	0.60(0.49)				
k12_fu	K	NV	77	0.79(0.41)				
k12_fz	K	NV	76	0.33(0.47)				
k12_l3	K	WV	77	0.48(0.50)	-0.24	0.18	0.81	0.65
k12_loc	K	ZV	77	0.70(0.46)	-1.41	0.19	1.04	0.40
k12_mat	K	ZV	77	0.49(0.50)	-0.31	0.18	0.95	0.51
k13_fo	K	NV	76	0.70(0.46)				
k13_fu	K	NV	77	0.83(0.38)				
k13_fz	K	NV	77	0.27(0.45)				
k13_l3	K	WV	77	0.35(0.48)	0.43	0.19	0.98	0.48
k13_loc	K	ZV	77	0.57(0.50)				
k13_mat	K	ZV	77	0.39(0.49)	0.22	0.19	0.87	0.57
k14_ad	K	NV	77	0.60(0.49)				
k14_fu	K	NV	77	0.48(0.50)				
k14_fz	K	NV	77	0.36(0.48)				
k14_l3	K	WV	77	0.64(0.48)	-1.09	0.19	0.99	0.45
k14_loc	K	ZV	77	0.32(0.47)	0.57	0.19	0.94	0.51
k14_mat	K	ZV	76	0.72(0.45)				
k15_ad	K	NV	77	0.60(0.49)				
k15_fo	K	NV	77	0.75(0.43)				
k15_fu	K	NV	77	0.13(0.34)	1.88	0.21	1.09	0.24
k15_l3	K	WV	77	0.41(0.49)	0.14	0.19	0.85	0.60
k15_loc	K	ZV	77	0.73(0.45)	-1.59	0.20	0.99	0.43
k15_mat	K	ZV	77	0.45(0.50)	-0.11	0.18	0.69	0.74
k16_ad	K	NV	77	0.73(0.45)				

k16_fu	K	NV	76	0.55(0.50)				
k16_fz	K	NV	77	0.18(0.39)				
k16_l3	K	WV	77	0.55(0.50)	-0.57	0.18	1.06	0.44
k16_loc	K	ZV	77	0.40(0.49)	0.16	0.18	1.01	0.45
k16_mat	K	ZV	77	0.42(0.50)	0.10	0.18	0.95	0.51
k17_ad	K	NV	77	0.48(0.50)				
k17_fu	K	NV	77	0.77(0.43)				
k17_fz	K	NV	77	0.14(0.35)	1.80	0.21	1.11	0.12
k17_l3	K	WV	77	0.39(0.49)	0.23	0.19	0.70	0.73
k17_loc	K	ZV	77	0.31(0.47)	0.65	0.19	0.91	0.53
k17_mat	K	ZV	77	0.70(0.46)	-1.40	0.19	0.87	0.56
k18_ad	K	NV	74	0.69(0.47)				
k18_fu	K	NV	76	0.58(0.50)				
k18_fz	K	NV	76	0.17(0.38)				
k18_l3	K	WV	77	0.52(0.50)	-0.40	0.19	1.07	0.41
k18_loc	K	ZV	77	0.50(0.50)	-0.33	0.18	0.95	0.54
k18_mat	K	ZV	77	0.42(0.50)	0.10	0.18	0.82	0.63
k19_ad	K	NV	75	0.76(0.43)				
k19_fu	K	NV	77	0.47(0.50)				
k19_fz	K	NV	77	0.17(0.38)				
k19_l3	K	WV	77	0.45(0.50)	-0.10	0.18	0.82	0.64
k19_loc	K	ZV	77	0.35(0.48)	0.44	1.39	1.04	0.42
k19_mat	K	ZV	77	0.38(0.49)	0.30	0.19	0.72	0.71

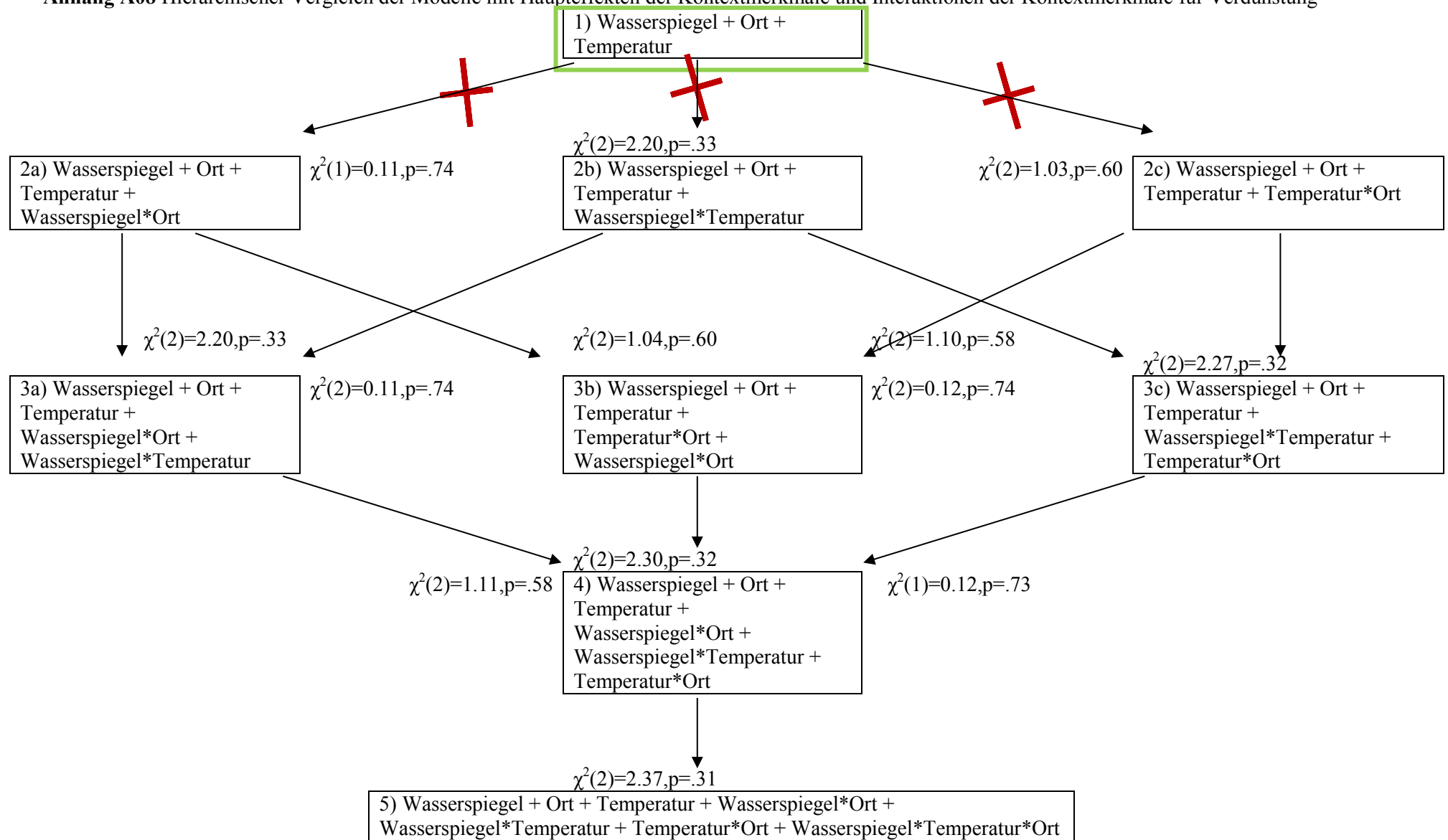
*Anmerkungen.* NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; A, Auftrieb; D, Dichte; K, Kondensation; V, Verdrängung; Vd, Verdunstung; <sup>a</sup>Die Werte entstammen der eindimensionalen Rasch-Skalierung. <sup>b</sup>Items, für die keine Rasch-Werte angegeben sind, wurden aufgrund ungenügender psychometrischer Qualität aus den Analysen ausgeschlossen.

**Anhang A07.3****Kennwerte der Items zu Dichte und Verdrängung aus Studie 2**

Item	Be- reich	Niveau	N	M(SD)	estimate <sup>a</sup>	error <sup>a</sup>	weighted MNSQ <sup>a</sup>	Trenn- schärfe <sup>a</sup>
d1	D	WV	73	0.59(0.50)	0.12	0.23	0.70	0.83
d10	D	WV	73	0.92(0.28)	-2.80	0.29	1.11	0.35
d11	D	WV	73	0.85(0.36)	-1.92	0.26	1.26	0.38
d12	D	WV	73	0.82(0.39)	-1.65	0.74	1.20	0.43
d2	D	WV	73	0.59(0.50)	0.12	0.23	0.87	0.76
d3 <sup>b</sup>	D	WV	73	0.27(0.45)				
d4	D	WV	73	0.58(0.50)	0.22	0.23	0.85	0.73
d5	D	WV	73	0.37(0.49)	1.66	0.24	1.08	0.62
d6	D	WV	73	0.29(0.46)	2.30	0.25	1.07	0.57
d7	D	WV	73	0.51(0.50)	0.69	0.23	0.99	0.71
d8	D	WV	73	0.42(0.50)	1.26	0.24	0.91	0.75
d9	D	WV	73	0.10(0.30)				
v1	V	WV	72	0.94(0.23)				
v2	V	WV	72	0.71(0.46)	0.82	0.25	1.09	0.74
v3	V	WV	73	0.97(0.16)	-3.08	0.33	1.17	0.22
v4	V	WV	73	0.68(0.47)	1.04	0.25	0.84	0.84
v5	V	WV	73	0.68(0.47)	1.04	0.25	1.11	0.72
v6	V	WV	73	0.78(0.42)	0.19	0.54	0.98	0.74

*Anmerkungen.* NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; A, Auftrieb; D, Dichte; V, Verdrängung; <sup>a</sup>Die Werte entstammen der jeweils eindimensionalen Rasch-Skalierung der Bereiche Dichte und Verdrängung. <sup>b</sup>Items, für die keine Rasch-Werte angegeben sind, wurden aufgrund ungenügender psychometrischer Qualität aus den Analysen ausgeschlossen.

**Anhang A08** Hierarchischer Vergleich der Modelle mit Haupteffekten der Kontextmerkmale und Interaktionen der Kontextmerkmale für Verdunstung



## Anhang A09

Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Verdunstung

Untersuchtes Kontext-merkmal		Modell nur Haupteffekte		Modell Haupteffekte und Interaktion	
		B	SE	B	SE
Wasser- spiegel	(Intercept)	-0.01	0.17	0.00	0.19
	ZV	-0.32*	0.13	-0.31	0.19
	WV	-0.41*	0.16	-0.44 <sup>+</sup>	0.23
	Wasserspiegel=nein	-.015	0.12	-0.16	0.19
	ZV*Wasserspiegel=nein			-0.01	0.27
	WV*Wasserspiegel=nei				
	n			0.07	0.33
	BIC	22744		22764	
$\chi^2$		$\chi^2(2)=0.07, p=.97$			
Temperatur	(Intercept)	-0.02	0.18	0.05	0.22
	ZV	-0.32*	0.13	-0.34	0.24
	WV	-0.41*	0.16	-0.71*	0.29
	Temperatur=Sonne	0.01	0.15	-0.12	0.24
	Temperatur=kalt	-0.17	0.15	-0.25	0.23
	ZV*Temperatur=Sonne			0.07	0.33
	WV*Temperatur=Sonne			0.46	0.40
	ZV*Temperatur=kalt			-0.02	0.32
	WV*Temperatur=kalt			0.41	0.40
	BIC	22753		22791	
$\chi^2$		$\chi^2(4)=1.87, p=.76$			
Ort	(Intercept)	-0.10	0.17	-0.07	0.18
	ZV	-0.31*	0.13	-0.37*	0.19
	WV	-0.40*	0.16	-0.42 <sup>+</sup>	0.23
	Ort=draußen	0.02	0.12	-0.04	0.19
	ZV*Ort=draußen			0.13	0.27
	WV*Ort=draußen			0.05	0.33
	BIC	22691		22695	
	$\chi^2$		$\chi^2(2)=0.23, p=.89$		

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen;

<sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001;

## Anhang A10

Ergebnisse zu den Interaktionen zwischen Kontextmerkmalen und angenommenem Niveau einer Aufgabe für Kondensation

Untersuchtes Kontext-merkmal		Modell nur Haupteffekte		Modell Haupteffekte und Interaktion	
		B	SE	B	SE
Ort	(Intercept)	-1.92***	0.48	-1.84***	0.51
	ZV	1.70***	0.39	1.75***	0.44
	WV	1.87***	0.40	1.55***	0.46
	Ort=draußen	-0.14	0.19	-0.46	0.83
	ZV*Ort=draußen			0.07	0.86
	WV*Ort=draußen			0.81	0.88
	BIC	4918		4932	
		$\chi^2$		$\chi^2(2)=3.52, p=.17$	
Temperatur	(Intercept)	-2.05***	0.49	-2.22***	0.60
	ZV	1.67***	0.38	1.90***	0.54
	WV	1.85***	0.40	1.93***	0.56
	Temperatur=kalt	0.19	0.19	0.52	0.72
	ZV*Temperatur=kalt			-0.47	0.76
	WV*Temperatur=kalt			-0.14	0.79
	BIC	4918		4934	
		$\chi^2$		$\chi^2(4)=0.84, p=.66$	
Konden-sationskern	(Intercept)	-1.96***	0.48	-1.88***	0.52
	ZV	1.65***	0.39	1.53***	0.44
	WV	1.84***	0.40	1.80***	0.46
	Kern=Dampf	0.05	0.22	-0.29	0.84
	ZV*Kern=Dampf			0.51	0.89
	WV*Kern=Dampf			0.11	0.93
	BIC	4919		4935	
		$\chi^2$		$\chi^2(2)=0.84, p=.66$	
Quelle	(Intercept)	-1.95***	0.47	-	-
	ZV	1.49***	0.38	-	-
	WV	1.68***	0.40	-	-
	Quelle=nein	0.35 <sup>+</sup>	0.19	-	-
	ZV*Quelle=nein			-	-
	WV*Quelle=nein			-	-
	BIC	4916		-	
		$\chi^2$		-	

Anmerkungen. NV, naive Vorstellungen; ZV, Zwischenvorstellungen; WV, wissenschaftliche Vorstellungen; Kern, Kondensationskern; <sup>+</sup>p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001;

## 9 Verzeichnis der elektronischen Anhänge

### 1\_ Studie 1- Querschnittstudie

- 1.01\_ Testheft 1 - Klasse 2.pdf<sup>14</sup>
- 1.02\_ Testheft 1 - Klasse 3 u 4.pdf
- 1.03\_ Testheft 2 - Klasse 2.pdf
- 1.04\_ Testheft 2 - Klasse 3 u 4.pdf
- 1.05\_ Testheft 3 - Klasse 2.pdf
- 1.06\_ Testheft 3 - Klasse 3 u 4.pdf
- 1.07\_ Testheft 4 - Klasse 2.pdf
- 1.08\_ Testheft 4 - Klasse 3 u 4.pdf
- 1.09\_ Testheft 5 - Klasse 2.pdf
- 1.1\_ Testheft 5 - Klasse 3 u 4.pdf
- 1.11\_ Testheft 6 - Klasse 2.pdf
- 1.12\_ Testheft 6 - Klasse 3 u 4.pdf

### 2\_ Studie 2- Experimentalstudie

- 2.01\_ Testheft Verdunstung.pdf
- 2.02\_ Testheft Kondensation.pdf
- 2.03\_ Testheft Dichte u Verdrängung.pdf

---

<sup>14</sup> Die elektronischen Anhänge sind auf dem beigegeführten Datenträger zu finden. Alle Testhefte in den elektronischen Anhängen sind in der Testleiterversion auf dem Datenträger vorhanden.

## **Plagiatserklärung**

Hiermit versichere ich, dass die vorliegende Arbeit mit dem Titel *Kontextmerkmale und die Bearbeitung von Aufgaben in einem Test naturwissenschaftlicher Kompetenz in der Grundschule* selbstständig verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

---

(Ort, Datum, Unterschrift)



## Lebenslauf











