

Cavelti, Martina Flurina

Entwicklung und Validierung eines Messinstruments zur Erfassung der Schülerkompetenzen im Bereich des wissenschaftlichen Skizzierens im Fach Chemie in der Sekundarstufe I

Berlin : Logos Verlag 2024, 248 S. - (Studien zum Physik- und Chemielernen; 377) - (Dissertation, Universität Duisburg-Essen, 2024)



Quellenangabe/ Reference:

Cavelti, Martina Flurina: Entwicklung und Validierung eines Messinstruments zur Erfassung der Schülerkompetenzen im Bereich des wissenschaftlichen Skizzierens im Fach Chemie in der Sekundarstufe I. Berlin : Logos Verlag 2024, 248 S. - (Studien zum Physik- und Chemielernen; 377) - (Dissertation, Universität Duisburg-Essen, 2024) - URN: urn:nbn:de:0111-pedocs-334726 - DOI: 10.25656/01:33472; 10.30819/5829

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-334726>

<https://doi.org/10.25656/01:33472>

in Kooperation mit / in cooperation with:



<https://www.logos-verlag.de>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Martina Flurina Cavelti

**Entwicklung und Validierung
eines Messinstruments zur Erfassung
der Schülerkompetenzen im Bereich
des wissenschaftlichen Skizzierens
im Fach Chemie in der Sekundarstufe I**



λογος

Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf und Mathias Ropohl

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

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

Martin Hopf und Mathias Ropohl

Studien zum Physik- und Chemielernen

Band 377

Martina Flurina Cavelti

**Entwicklung und Validierung eines
Messinstruments zur Erfassung der
Schülerkompetenzen im Bereich des
wissenschaftlichen Skizzierens im Fach
Chemie in der Sekundarstufe I**

Logos Verlag Berlin



Studien zum Physik- und Chemielernen

Martin Hopf und Mathias Ropohl [Hrsg.]

Bibliografische Information der Deutschen Nationalbibliothek

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



Dieses Werk ist lizenziert unter der Creative Commons Attribution 4.0 Lizenz CC BY (<https://creativecommons.org/licenses/by/4.0/>). Die Bedingungen der Creative-Commons-Lizenz gelten nur für Originalmaterial. Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit Quellenangabe) wie z.B. Schaubilder, Abbildungen, Fotos und Textauszüge erfordert ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.

Logos Verlag Berlin GmbH 2024

ISBN 978-3-8325-5829-1

ISSN 1614-8967

DOI 10.30819/5829

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

Tel.: +49 (0)30 / 42 85 10 90

Fax: +49 (0)30 / 42 85 10 92

<https://www.logos-verlag.de>

Entwicklung und Validierung eines Messinstruments zur
Erfassung der Schülerkompetenzen im Bereich
des wissenschaftlichen Skizzierens
im Fach Chemie in der Sekundarstufe I

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften

– Dr. rer. nat. –

vorgelegt von

Martina Flurina Cavelti

Fakultät Chemie der Universität Duisburg-Essen
Pädagogische Hochschule Zürich

2024

Die vorliegende Arbeit wurde im Zeitraum von August 2018 bis Februar 2024 in der Didaktik der Chemie der Universität Duisburg-Essen unter der Leitung von Prof. Dr. Maik Walpuski und an der Pädagogischen Hochschule Zürich unter der Leitung von Prof. Dr. Christoph Gut durchgeführt.

Gutachter: Prof. Dr. Maik Walpuski
Prof. Dr. Christoph Gut

Disputation: 19. Juni 2024

Vorsitzender: Prof. Dr. Micheal Giese

Diese Publikation wurde finanziert durch den Open-Access-Fonds (iOAF) der Pädagogischen Hochschule Zürich

Das Äußerste geben.
Immer mit dem Herzen arbeiten
und mit ganzem Herzen,
ob es sich darum handelt,
ein Raumschiff
zu den Sternen zu führen
oder einen
einfachen Punkt mit dem Bleistift
zu zeichnen.

Don Helder Camara

Inhaltsverzeichnis

ABBILDUNGSVERZEICHNIS

TABELLENVERZEICHNIS

FORMELVERZEICHNIS

1	ABSTRACT	1
2	EINLEITUNG	5
3	THEORETISCHER HINTERGRUND	9
3.1	KOMPETENZ	9
3.1.1	<i>Kompetenzbegriff</i>	9
3.1.2	<i>Experimentelle Kompetenz</i>	11
3.2	ERKENNTNISGEWINNUNG	14
3.2.1	<i>Naturwissenschaftliche Arbeitsweisen und wissenschaftliches Denken</i>	15
3.2.2	<i>Methoden der Erkenntnisgewinnung</i>	16
3.3	REPRÄSENTATIONSKOMPETENZ	25
3.4	SKIZZIEREN	29
3.4.1	<i>Beschreibung Skizzieren, Einordnung wissenschaftliches Skizzieren</i>	29
3.4.2	<i>Funktionen des Skizzierens</i>	32
3.4.3	<i>Fazit Skizzieren</i>	41
3.5	KOMPETENZMODELLE	42
3.5.1	<i>Kompetenzmodelle in Bezug auf das wissenschaftliche Skizzieren</i>	44
3.5.2	<i>Qualitätsmerkmale wissenschaftlichen Skizzierens</i>	46
4	ÜBERLEGUNGEN ZUR MODELLIERUNG DER KOMPETENZ DES WISSENSCHAFTLICHEN SKIZZIERENS	51
4.1	MODELLIERUNG IN BEZUG ZUR ERKENNTNISGEWINNUNG	53
4.2	MODELLIERUNG IN BEZUG ZUR EXPERIMENTELLEN KOMPETENZ	55
4.3	MODELLIERUNG IN BEZUG ZUR REPRÄSENTATIONSKOMPETENZ (RC)	56
4.4	MODELLIERUNG IN BEZUG AUF EXTERNE FAKTOREN	57
4.5	FAZIT MODELLIERUNG	60
5	FORSCHUNGSZIEL UND FORSCHUNGSFRAGEN	63
6	METHODEN	69
6.1	ENTWICKLUNG DES KOMPETENZSTRUKTURMODELLS	70
6.1.1	<i>Kompetenzstruktur der Aufgabentypen</i>	70
6.1.2	<i>Innere Differenzierung der Kompetenzstruktur</i>	72
6.1.3	<i>Progression und Einbettung der Qualitätsmerkmale ins Kompetenzstrukturmodell</i>	74
6.2	PROGRESSIONSMODELLE DES WISSENSCHAFTLICHEN SKIZZIERENS	75
6.3	ENTWICKLUNG AUFGABENSET	82
6.3.1	<i>Aufgabenstruktur und Aufgabenstamm (Vignette)</i>	85
6.3.2	<i>Video-Vignetten</i>	86
6.3.3	<i>Theorieblatt „Regeln des wissenschaftlichen Skizzierens“</i>	87
6.3.4	<i>Entwicklung Fach- und Methodenwissenstest</i>	87
6.3.5	<i>Instrumente der externen Validierung</i>	89
6.4	ERHEBUNG	90
6.4.1	<i>Stichprobe und Testdesign</i>	90
6.4.2	<i>Vorilotstudie</i>	90

6.4.3	<i>Weiterentwicklung Aufgabenset nach Evaluation Vorpilotierung</i>	93
6.4.4	<i>Pilotstudie</i>	96
6.4.5	<i>Hauptstudie</i>	97
6.5	KODIERUNG	99
6.5.1	<i>Entwicklung des Kodiermanual</i>	99
6.5.2	<i>Festlegung der Schwellenwerte</i>	100
6.5.3	<i>Aufgabentypabhängige Unterschiede der Items und der Indikatoren</i>	108
6.5.4	<i>Kontextabhängige Unterschiede der Items und der Indikatoren</i>	109
6.6	RELIABILITÄT DER KODIERUNG	110
6.6.1	<i>Prozentuale Übereinstimmung</i>	111
6.6.2	<i>Cohens Kappa</i>	112
6.6.3	<i>Gwet's AC1</i>	112
6.6.4	<i>Interrater-Reliabilität</i>	113
6.6.5	<i>Auffällige Indikatoren</i>	117
6.7	VALIDIERUNG	119
6.7.1	<i>Validierung der Qualitätsmerkmale des Kompetenzstrukturmodell</i>	120
6.7.2	<i>Ökologische Validierung des Aufgabensets</i>	120
6.7.3	<i>Strukturelle Validierung</i>	121
6.7.4	<i>Generalisierbarkeit</i>	121
6.7.5	<i>Externe Validierung</i>	121
6.8	PROBABILISTISCHE TESTTHEORIE	122
6.8.1	<i>Das Rasch-Modell</i>	122
6.8.2	<i>Das „Partial Credit Model“</i>	124
6.9	REGRESSIONSANALYSE	126
6.10	KORRELATIONSANALYSE	128
6.11	STRUKTURGLEICHUNGSMODELL	129
7	RESULTATE UND INTERPRETATIONEN	131
7.1	DATENAUFBEREITUNG	131
7.2	VALIDIERUNG DER QUALITÄTSMERKMALE DES KOMPETENZSTRUKTURMODELLS (F1)	133
7.3	ÖKOLOGISCHE VALIDIERUNG (F2)	136
7.4	TESTGÜTE, MESSINVARIANZ UND INTERPRETATION DER TESTWERTE (F3)	138
7.4.1	<i>Testgüte der Skala der Kompetenz des wissenschaftlichen Skizzierens</i>	138
7.4.2	<i>Schwierigkeitserzeugende Objektmerkmale</i>	142
7.4.3	<i>Aufgabentypabhängige Unterschiede der Itemschwierigkeiten</i>	142
7.4.4	<i>Kontextabhängigkeit der Itemschwierigkeiten</i>	143
7.4.5	<i>Schwellenwerte der Itemschwierigkeiten</i>	144
7.4.6	<i>Messinvarianz des Messinstruments</i>	152
7.5	EXTERNE VALIDIERUNG (F4-F6)	159
7.5.1	<i>Kognition, Strategiewissen zum Experimentieren und Lesefähigkeit (F4)</i>	160
7.5.2	<i>Fach- und Methodenwissen (F5)</i>	165
7.5.3	<i>externe Validierung einzelner Aufgabentypen (F6)</i>	169
8	ZUSAMMENFASSENDE DISKUSSION UND AUSBLICK	173
8.1	ZUSAMMENFASSENDE DISKUSSION	173
8.2	AUSBLICK	190

9	LITERATURVERZEICHNIS	193
10	DANKSAGUNG	205
11	ANHANG	207
11.1	FRAGENBOGEN ÖKOLOGISCHE VALIDIERUNG	207
11.2	FRAGEBOGEN INHALTLICHE VALIDIERUNG BEZÜGLICH DER QUALITÄTSMERKMALE DES WISSENSCHAFTLICHEN SKIZZIERENS	210
11.3	THEORIEBLATT	212
11.4	AUFGABENSET	213
11.4.1	<i>Aufgabenstellung Item BR_V (Vergleich, Brausetablette)</i>	213
11.5	KODIERMANUAL	219
11.5.1	<i>Kodiermanual Item BR_V (Vergleich, Brausetablette)</i>	219
11.6	INDIKATOREN	232
11.6.1	<i>Indikatoren BR_B</i>	232
11.6.2	<i>Indikatoren EI_B</i>	233
11.6.3	<i>Indikatoren SB_B</i>	234
11.6.4	<i>Indikatoren BR_V</i>	235
11.6.5	<i>Indikatoren EI_V</i>	236
11.6.6	<i>Indikatoren SB_V</i>	238
11.6.7	<i>Indikatoren BR_E</i>	239
11.6.8	<i>Indikatoren EI_E</i>	240
11.6.9	<i>Indikatoren SB_E</i>	242
11.7	INTERRATER AUSWERTUNG (ÜBEREINSTIMMUNG/COHENS KAPPA/GWET AC1)	243

Abbildungsverzeichnis

Abb. 1 Formen von Dokumentationen	30
Abb. 2 entwickelndes Verständnis für das Skizzendesign in den Skizzentypen nach dem Framework von Hope (2008)	45
Abb. 3 Vergleich der Niveaus von Hope's Framework den Niveaus des wissenschaftlichen Skizzierens	82
Abb. 4 Beispiel einer Vignette	86
Abb. 5 Konstruktion des individuellen Schülercodes	90
Abb. 6 Schülerskizzen des Kontexts „Eis schmelzen“ im Aufgabentyp „Beobachtung“ (EI_B)	116
Abb. 7 Schülerskizzen des Kontextes „Brausetablette“ im Aufgabentyp „Vergleich“ (BR_V)	117
Abb. 8 Validierung der Qualitätsmerkmale wie auch des Aufgabentyps: „unverzichtbar“ = 4, „eher ja“ = 3, „eher nein“ = 2, „gar nicht“ = 1	133
Abb. 9 ökologische Validierung	137
Abb. 10 Wright-Map der Itemschwierigkeit, aufgeteilt nach Aufgabentypen „Wissenschaftlichkeit der Skizze“ mit den aufgabentypspezifischen Kategorien links (x = 2.4 Fälle), ohne aufgabentypspezifische Kategorie rechts (x = 2.6 Fälle)	140
Abb. 11 Wright-Map der Itemschwierigkeiten (Schwellenwerte, getrennt nach Aufgabentyp und Item) (x = 2.6 Fälle)	146
Abb. 12 Schülerskizze des Aufgabentyps „Experimentieren“ im Kontext „Brausetablette“ (Item BR_E)	147
Abb. 13 Schülerskizze des Aufgabentyps „Beobachtung“ im Kontext „Eis schmelzen“ (Item EI_B)	148
Abb. 14 Schülerskizze des Aufgabentyps „Vergleich“ im Kontext „Säuren und Basen“ (Item SB_V)	149
Abb. 15 Schülerskizze des Aufgabentyps „Vergleich“ im Kontext „Brausetablette“ (Item BR_V)	150
Abb. 16 Schülerskizze des Aufgabentyps „Beobachtung“ im Kontext „Säuren und Basen“ (Item SB_B)	151
Abb. 17 Strukturgleichungsmodell (Mplus) mit KFT als übergeordneter Faktor als Modell dargestellt, wobei * = $p < 0.11$, ** = $p < 0.001$	162

Abb. 18 Strukturgleichungsmodell (Mplus) mit allen Faktoren gleich gewichtet als Modell dargestellt, wobei * * = $p < 0.11$, ** = $p < 0.001$	163
Abb. 19 Darstellung der Varianteneinteilung der erbrachten Performanz in Fach- und Methodenwissen und dem wissenschaftlichen Skizzieren.....	165

Tabellenverzeichnis

Tab. 1 Gegenüberstellung konzeptueller Charakteristika des Kompetenz- und Intelligenzbegriffs (Hartig & Klieme, 2006).....	10
Tab. 2 Struktur der Erkenntnisgewinnung in der Chemie nach Nehring et al. (2016)	16
Tab. 3 Vergleichsdimension (Verhältnis der experimentellen Kompetenz, der Erkenntnisgewinnung und der RC zum wissenschaftlichen Skizzieren)	53
Tab. 4 Grundanforderungen und resultierende Erkenntnisgewinnung	54
Tab. 5 Forschungsdesign mit Aufgabentypen.....	64
Tab. 6 Kategorien und Aufgabenstellung der Aufgabentypen.....	72
Tab. 7 Beschreibung der Facetten der Kategorie „Wissenschaftlichkeit der Skizze“	73
Tab. 8 Literaturbezug der Inhalte der Facetten der Kategorie „Wissenschaftlichkeit der Skizze“	75
Tab. 9 Progressionsmodell des Aufgabentyps „Beobachtung“	77
Tab. 10 Progressionsmodell des Aufgabentyps „Vergleich“	78
Tab. 11 Progressionsmodell des Aufgabentyps „Experimentieren“	80
Tab. 12 Kontexte und Aufgabenstellungen der Vorpilotierung	83
Tab. 13 Fragenstruktur des Aufgabentyps „Beobachtung“	88
Tab. 14 Fragenstruktur des Aufgabentyps „Vergleich“	88
Tab. 15 Fragenstruktur des Aufgabentyps „Experimentieren“	89
Tab. 16 Units in der Vorpilotstudie.....	91
Tab. 17 Rotationsschema Vorpilotstudie	91
Tab. 18 Erhebungsmanual Vorpilotierung für ein Item	92
Tab. 19 Schwierigkeiten bei den Experimenten und Reaktionen oder den Skizzen.	93
Tab. 20 Items der Pilot- und Hauptstudie.....	95
Tab. 21 Abkürzung der Item-Bezeichnungen	95
Tab. 22 Zusammensetzung Units Pilotstudie	97
Tab. 23 Rotationsschema Pilotstudie	97
Tab. 24 Rotationsschema Hauptstudie	98
Tab. 25 Überblick über die Begrifflichkeiten	100
Tab. 26 Schwellenwerte: Kontexte „Brausetablette und „Säuren und Basen“, Aufgabentyp „Beobachtung“	102

Tab. 27 gewichtete Prozente der Erfüllung der Facette und Rang BR_B	103
Tab. 28 gewichtete Prozente der Erfüllung der Facette und Rang EI_V	103
Tab. 29 Abweichungspunkte des Niveaus unter Ausschluss einer Facette des Kontext Brausetablette und des Items EI_V	104
Tab. 30 Schwellenwerte: Kontext „Brausetablette“, Aufgabentyp „Vergleich“	106
Tab. 31 Schwellenwerte: Kontext „Eis schmelzen“, Aufgabentyp „Vergleich“	106
Tab. 32 Schwellenwerte: Kontext „Säuren und Basen“, Aufgabentyp „Vergleich“	107
Tab. 33 Schwellenwerte: Aufgabentyp „Experimentieren“, alle drei Kontexte	107
Tab. 34 Gemeinsamkeiten und Unterschiede zwischen den Aufgabentypen Beobachtung/ Vergleich und Experimentieren	108
Tab. 35 Objektmerkmale, die zu Indikatoren der Facette „fachliche Korrektheit“ führen.....	109
Tab. 36 Prozentsatz verglichener Testhefte	111
Tab. 37 Zusammenstellung niedrig-inferenter Indikatoren.....	113
Tab. 38 Übereinstimmungen aller verglichenen Kriterien, aufgeschlüsselt nach Aufgabentyp und Kontext	114
Tab. 39 Indikatoren mit ungenügender Objektivität	115
Tab. 40 Zusammenstellung Indikatoren mit niedrigen Mittelwerten des Aufgabentyps „Experimentieren“	118
Tab. 41 Einstufung der Korrelationskoeffizienten (Zöfel, 2003).....	129
Tab. 42 Item-Kennwerte der neun Items des NAWs der Rasch-Analyse	132
Tab. 43 Validierung der Qualitätsmerkmale und des Aufgabentyps	134
Tab. 44 Kriterien und Qualitätsmerkmale Nennungen einzelner Fachdidaktik Expertinnen und -experten	135
Tab. 45 ökologische Validierung: Kontext, Aufgabentyp und Mittelwert.....	138
Tab. 46 Item-Kennwerte der neun Items des wissenschaftlichen Skizzierens der Rasch-Analyse.....	139
Tab. 47 Item-Kennwerte der neun Items der Rasch-Analyse	139
Tab. 48 Item-Kennwerte der neun Items der Rasch-Analyse (Abb. 11).....	144
Tab. 49 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich des Geschlechts	153
Tab. 50 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich der Jahrgangsstufe	154

Tab. 51 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich der Anzahl der gesprochenen Sprachen	156
Tab. 52 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich des Leistungsniveaus.....	157
Tab. 53 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich der Personenfähigkeit	158
Tab. 54 Resultate externe Variablen	160
Tab. 55 Korrelationskoeffizienten zwischen der Personenfähigkeit des wissenschaftlichen Skizzierens und der Lesefähigkeit	161
Tab. 56 Prädiktoren für die Personenfähigkeit.....	161
Tab. 57 Kennwerte für das Strukturgleichungsmodell (Mplus).....	162
Tab. 58 erbrachte Leistungen im wissenschaftlichen Skizzieren und Multiple-Select-Test zum Fach- und Methodenwissen in Prozent.....	166
Tab. 59 Kendall-Tau-Test zwischen den Kompetenzen im wissenschaftlichen Skizzieren und den Resultaten des Multiple-Select-Tests zum Fach- und Methodenwissen.....	168
Tab. 60 Item-Kennwerte der 2D-Rasch-Analyse	169
Tab. 61 Prädiktoren für die Personenfähigkeit, 2D Betrachtung für die Aufgabentypen „Beobachtung“ und „Vergleich“	170
Tab. 62 Prädiktoren für die Personenfähigkeit, 2D Betrachtung für den Aufgabentyp „Experimentieren“	170

Formelverzeichnis

Formel 1 Rasch-Modell-Gleichung (Koller et al., 2012)	123
Formel 2 Effektstärke der Regressionsanalyse	128

1 Abstract

Skizzieren ist wichtig und relevant für den naturwissenschaftlichen Unterricht. Es stellt eine Kompetenz dar, wird aber auch als unterstützende Methode beispielsweise zur Lernförderung und als Messinstrument zur Erfassung von anderen Kompetenzen genutzt. Das Skizzieren kommt also in unterschiedlichen Kontexten zur Anwendung. Während experimenteller Handlungen sollten Lernende ihre Vorgehensweise sowie ihre Ergebnisse und Beobachtungen verbal-sprachlich protokollieren und skizzierend festhalten können. Das sogenannte „*wissenschaftliche Skizzieren*“ bezeichnet die Kompetenz, während des Experimentierprozesses Skizzen anzufertigen, die die Grundanforderungen der Eindeutigkeit, Idealisierung, Vollständigkeit, Abstraktion und Korrektheit erfüllen. Dadurch können Teilprozesse des experimentellen Handelns, wie die Entwicklung der Fragestellung, die Planung und Durchführung des Experiments, Messungen und Beobachtungen sowie das Ziehen von Schlussfolgerungen, festgehalten und aus diesen *wissenschaftlichen Skizzen* Erkenntnisse gewonnen werden. Da anstelle des Schreibens skizziert wird, sinken die verbal-sprachlichen Anforderungen an Schülerinnen und Schüler. Eine *wissenschaftliche Skizze* ist die förderliche Basis für Austausch und Kommunikation wie auch für die Erkenntnisgewinnung, da durch eine *wissenschaftliche Skizze*, die die Grundanforderungen abdeckt, Wesentliches erkannt und daraus Erkenntnis gewonnen werden kann.

Zum *wissenschaftlichen Skizzieren* als Schülerkompetenz existiert in der Literatur bislang kein spezifisches Modell, das die Struktur oder die Stufung einer entsprechenden Kompetenz beschreibt und somit eine Möglichkeit zur Messbarkeit dieser Kompetenz böte.

Die ersten Forschungsfragen befassen sich mit der inhaltlichen Validierung der Qualitätsmerkmale eines Kompetenzstrukturmodells, mit der ökologischen Validierung und der Überprüfung der Gütekriterien Objektivität, Reliabilität und Validität sowie außerdem mit der Generalisierbarkeit des Messinstruments, welches auf der Kompetenz des *wissenschaftlichen Skizzierens* entwickelt wurde und die Schülerkompetenz *wissenschaftliches Skizzieren* auf der Basis eines a-priori-Kompetenzmodells misst. Im Fokus steht die Fähigkeit der Lernenden, *wissenschaftliche Skizzen* während der Planungs-, Durchführungs- und Beobachtungsphase des experimentellen Handelns in

der Sekundarstufe I im Fach Chemie anzufertigen. Weitere Forschungsfragen behandeln die externe Validität des Messinstruments.

Im Rahmen dieser Arbeit entstanden ein literaturbasiertes Kompetenzmodell mit Progressionsmodell und ein Messinstrument für die Kompetenz des *wissenschaftlichen Skizzierens*. Das Messinstrument umfasst neun Items, die aus drei Kontexten (Brausetablette, Eis schmelzen, Säuren und Basen) stammen, und aus drei Aufgabentypen (Beobachtung, Vergleich und Experimentieren) bestehen. Die Validität des Kompetenzstrukturmodells und des Tests wurde zum einen inhaltlich und ökologisch mittels Expertenbefragung und zum anderen bezüglich Generalisierbarkeit und Testgüte überprüft. Die erhobenen Daten des *wissenschaftlichen Skizzierens* wurden mit Rasch-Analysen ausgewertet und in Bezug auf die Forschungsfragen analysiert.

Die Daten zur Überprüfung der Messinvarianz des Tests wurden nach Geschlecht, Jahrgangsstufe, Anzahl gesprochener Sprachen und Fähigkeiten im *wissenschaftlichen Skizzieren* mit DIF-Analysen differenziert ausgewertet, sodass generalisierbare Schlüsse gezogen werden konnten. Für die externe Validität wurden die Zusammenhänge für die Faktoren Kognition (KFT), Lesefähigkeit (LGVT), Strategiewissen zum Experimentieren (NAW) sowie Fach- und Methodenwissen mit der Kompetenz des *wissenschaftlichen Skizzierens* untersucht.

Das Progressionsmodell wie auch das Testinstrument zeigten sich inhaltlich und ökologisch valide und lassen Verallgemeinerungen zu. Das Testinstrument weist Testgüte auf und ist reliabel und messinvariant in Bezug auf das Geschlecht, die Jahrgangsstufe, die Anzahl gesprochener Sprachen und die Personenfähigkeiten im *wissenschaftlichen Skizzieren*. Die Items lassen sich auch von schwächeren Lernenden bewältigen. Die Kompetenz des *wissenschaftlichen Skizzierens* kann in vier Niveaus unterteilt werden.

Über Zusammenhänge mit externen Faktoren lassen sich folgende Aussagen machen: Die Kompetenz des *wissenschaftlichen Skizzierens* zeigt einen signifikant positiven, starken Zusammenhang mit der Kognition, einen signifikanten, wenn auch schwächeren Zusammenhang mit dem Strategiewissen zum Experimentieren, und sie ist unabhängig von der Lesefähigkeit. Das Fach- und Methodenwissen zeigt bei den Aufgabentypen „Beobachtung“ und „Vergleich“ einen positiven Zusammenhang zum *wissenschaftlichen Skizzieren*, wohingegen beim Aufgabentyp „Experimentieren“ ein ne-

gativer Zusammenhang zum Methodenwissen der Variablenkontrollstrategie erkannt wurde.

Das neu entwickelte und inhaltlich, ökologisch und extern validierte Messinstrument eignet sich zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* in der Sekundarstufe I.

2 Einleitung

Experimentelle Kompetenzen stellen im Rahmen von Bildungsstandards ein zentrales Bildungsziel in der Schweiz und auch in Deutschland dar (EDK Erziehungsdirektoren-Konferenz, 2011; Kultusministerkonferenz, 2005a, 2005b, 2005c). Experimentieren hat in der Schweiz aus zwei Gründen mehr Gewicht erlangt: einerseits durch die interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) mit den sogenannten Handlungsaspekten im HarmoS-Kompetenz-Modell der Naturwissenschaften (EDK Erziehungsdirektoren-Konferenz, 2011), andererseits durch die Einführung des neuen schweizweit verbindlichen Lehrplans 21 (LP21) (D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz, 2016a) mit ab dem Jahre 2017 geltenden Basisstandards und der damit verbundenen Kompetenzorientierung. Auch in Deutschland hat sich der Stellenwert des Experimentierens durch Setzen von Bildungsstandards vergrößert (Kultusministerkonferenz, 2005c, 2005b, 2005a). Die experimentellen Handlungen erfordern auch Fähigkeiten des Dokumentierens, wodurch dem Dokumentieren eine gewichtige Funktion zukommt. Handlung, Phänomene, Planung und Ergebnisse während des Experimentierens können sowohl verbal-sprachlich protokolliert und als auch zeichnerisch dargestellt werden.

Zur Sicherung der Qualität und der Durchlässigkeit des schweizerischen Bildungssystems innerhalb der Schweiz legte HarmoS verbindliche Grundkompetenzen in den Fächern fest, wie beispielsweise im Handlungsaspekt „Fragen und Untersuchen“ des Bereichs Natur und Technik: Lernenden sollen u. a. „Ergebnisse aus Erkundungen, Untersuchungen und Experimente in verschiedenen Formen - insbesondere als Skizze, Bericht, Tabelle, Diagramm, Plan - einfach darstellen und kommentieren“ (EDK Erziehungsdirektoren-Konferenz, 2011, S. 33) können. Des Weiteren werden für die Handlungsaspekte „Ordnen, Strukturieren und Modellieren“ und „Mitteilen und Austauschen“ explizit Darstellungen in Form von Zeichnungen und Grafiken gefordert. Der Lehrplan 21 postuliert in den didaktischen Hinweisen (NMG/NT), dass Schülerinnen und Schüler „Beobachtungen und Gedanken sprachlich präzise ausdrücken und sie sich anderen mitteilen können.“ (D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz, 2016a, S. 233). Im Kompetenzbereich NT.3 (chemische Reaktionen erforschen) wird gefordert, dass Schülerinnen und Schüler u. a. angeleitet Reaktionen mit Sauerstoff durchführen, protokollieren, Fragen stellen, Vermutungen

formulieren und diese experimentell überprüfen können. Dadurch wird das Dokumentieren bereits umfassend in den Forschungszyklus eingebettet.

In den deutschen Bildungsstandards für das Fach Chemie sind Kommunikation und Erkenntnisgewinnung neben Fachwissen und Bewertung gleichwertige Kompetenzbereiche (Kultusministerkonferenz, 2005b). Beim Vergleich der Standards Kommunikation und Erkenntnisgewinnung sind Überlappungen erkennbar, die Standards sind nicht trennscharf voneinander formuliert (Kobow, 2015). Der Bereich Kommunikation ist in drei Aspekte aufgeteilt: Sprache/Fachsprache, Darstellungsformen und Adressatenbezug/Sachbezug. Zu den Darstellungsformen zählen neben Text auch Bilder, Zeichnungen, Tabellen, Diagramme, Formelzeichen und Gleichungen (Kobow, 2015). Die deutschen Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss (Kultusministerkonferenz, 2005b) fordern im Kompetenzbereich Erkenntnisgewinnung (experimentelle und andere Untersuchungsmethoden sowie Modelle nutzen) unter dem Standard E3, dass Lernende in der Lage sein müssen, experimentelle und andere Untersuchungen zu protokollieren. Im Kompetenzbereich Kommunikation (Informationen sach- und fachbezogen erschließen und austauschen) ist Dokumentieren in den Standards K4, K6 und K7 aufgeführt, wobei K4 für das Beschreiben, Veranschaulichen oder Erklären chemischer Sachverhalte unter Verwendung der Fachsprache und/oder mit Hilfe von Modellen und Darstellungen. steht, K6 das Protokollieren des Verlaufs und der Ergebnisse von Untersuchungen und Diskussionen in angemessener Form behandelt und K7 sich dem situationsgerechten und adressatenbezogenen Dokumentieren und Präsentieren des Verlaufs und der Arbeitsergebnisse) widmet. Auch in den Bildungszielen der USA wird Dokumentieren in diesem Sinne aufgeführt mit „Integrate quantitative or technical information expressed in words in a text with a version of that information expressed visually (e.g., in a flowchart, diagram, model, graph, or table)“ aufgeführt (NGA Center for Best Practices (NGA Center), Council of Chief State School Officers (CCSSO), 2010).

Protokollieren und Dokumentieren ist nach Süselbeck (2003) unbeliebt: Im Berufsleben allgemein, aber auch im Unterricht, und dies über alle Jahrgangsstufen hinweg (Kraus & Stehlik, 2008). Im naturwissenschaftlichen Unterricht ist das Dokumentieren von Experimenten vor allem wegen fehlender Selbstbestimmung im Dokumentierprozess unbeliebt (Bergeler, 2009). Nach Leisen (2003) führen Lernende lieber Experimente durch, als diese zu dokumentieren.

Die Unbeliebtheit des Dokumentierens, die Forderung der Bildungsstandards verschiedener Länder (D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz, 2016b; EDK Erziehungsdirektoren-Konferenz, 2011; Kultusministerkonferenz, 2005a, 2005b, 2005c; NGA Center for Best Practices (NGA Center), Council of Chief State School Officers (CCSSO), 2010) wie auch die Möglichkeit sprachentlastenden Dokumentierens in Form von Skizzen bilden die Grundlage für die Motivation zu dieser Arbeit. Ebenso zentral für die Stellenwert dieser Arbeit ist, dass *Skizzieren* sowohl aus wissenschaftlicher Sicht, z.B. als Repräsentationsfunktion (Ainsworth, 1999), als auch für die Schulpraxis als Handlungskompetenz relevant ist, da Experimentierprozesse häufig nicht allein verbal dargestellt werden können (Kozma et al., 2000). Außerdem arbeitet man in den Naturwissenschaften oft mit nicht direkt zugänglichen Aspekten und Phänomenen, die zu klein (z.B. Atome) oder zu groß (z.B. Galaxien) sind, weil sie in Zeiteinheiten stattfinden, die schwer wahrnehmbar sind (z.B. sehr schnelle chemische Reaktionen oder sehr langsame Plattentektonik) oder die zu komplex sind (Pande & Chandrasekharan, 2017). Durch Skizzen wird es möglich, diese Phänomene zu verstehen, zu analysieren und daraus Schlüsse zu ziehen. Skizzen bieten dadurch eine förderliche Basis für die Erkenntnisgewinnung. Gerade diesen Erkenntnisgewinnungsprozessen wird im naturwissenschaftlichen Unterricht eine hohe Wichtigkeit beigemessen.

Wissenschaftliches Skizzieren stellt also eine Teilfähigkeit dar, beim Dokumentieren objektive Skizzen eines Experimentierprozesses zu erstellen. Das *wissenschaftliche Skizzieren* erfüllt vielseitige und wichtige Zwecke beim Experimentieren: Schülerinnen und Schüler visualisieren ihre Vorgehensweise beim Experimentieren und halten Phänomene, Ergebnisse von Experimentierplanung, Beobachtungen und die Durchführung der Experimente fest. In *wissenschaftlichen Skizzen* werden Evidenzen dokumentiert. Sie stellen daher eine Repräsentationsform dar, die eine Basis für Kommunikation und Austausch darstellt.

In Bezug auf die Schülerkompetenz *wissenschaftliches Skizzieren* findet sich in der Literatur kein spezifisches Modell, welches die Struktur oder die Stufung einer entsprechenden Kompetenz beschreibt und somit eine Möglichkeit zur Messbarkeit dieser Kompetenz bieten könnte. Ziel dieses Dissertationsprojektes ist es, ein Messinstrument zu entwickeln und validieren, das die Leistung von Schülerinnen und Schülern in der Kompetenz des *wissenschaftlichen Skizzierens* erfasst. Die Entwicklung

und Durchführung Projektes fand in Kooperation der Pädagogischen Hochschule Zürich mit der Uni Duisburg-Essen statt.

3 Theoretischer Hintergrund

Für die Definition und Abgrenzung der Kompetenz des *wissenschaftlichen Skizzierens* bedarf es einer allgemeinen Definition des Begriffs Kompetenz wie auch einer definierten und beschreibenden Umgebung. *Wissenschaftliches Skizzieren* findet in experimentellen Settings statt und kann für die Erkenntnisgewinnung als förderliche Basis dienen. Eine Beleuchtung der experimentellen Kompetenz, wie der Erkenntnisgewinnung ist unabdingbar. Dabei werden die unterschiedlichen Ansätze der Experimentellen Kompetenz (Problemtyp- und Teilprozessansatz) wie auch die experimentelle Kompetenz nach Nehring (2016), bei der *Naturwissenschaftliche Arbeitsweise* und *wissenschaftliches Denken* matrixartig verknüpft sind, betrachtet. Ebenso zentral ist es, die Methoden der Erkenntnisgewinnung im Allgemeinen und spezifisch für die Anwendungsbereiche des *wissenschaftlichen Skizzierens* zu beleuchten. Neben der experimentellen Kompetenz spielt die Kompetenz im Umgang mit Repräsentationen, die *Repräsentationskompetenz RC*, für die theoretische Verordnung des *wissenschaftlichen Skizzierens* eine wichtige Rolle, da es sich bei Skizzen um eine Form von Repräsentationen handelt.

3.1 Kompetenz

3.1.1 Kompetenzbegriff

In der Wissenschaft nutzt man oftmals die Definition von Kompetenz nach Weinert (2001b), die im Auftrag der OECD (Organisation for Economic Co-operation and Development) erstellt wurde. Für die Kompetenzmessung in der Schulpraxis definiert Weinert (2001b) Kompetenzen wie folgt: „[...] Kompetenzen [sind] 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“. Zentrale Aspekte, die erlernt werden müssen, stellen dabei kognitive Fähigkeiten und Fertigkeiten dar (Klieme, 2004).

Kompetenzen sind themenspezifisch, das heißt sie gelten für einen bestimmten Bereich von Kontexten und Situationen. Bei Kompetenzen handelt es sich um erworbene Fähigkeiten und Fertigkeiten, die es ermöglichen, eine spezifische Aufgaben- oder

Problemstellung zu lösen (Hartig & Klieme, 2006). Trotz einer weitgefassten Definition des Kompetenzbegriffs wird oftmals aus pragmatischen Gründen bei der Überprüfung der Kompetenz, eine Beschränkung auf den kognitiven Bereich vollzogen, was eine Eingrenzung zur Folge hat (Klieme, 2004). Deshalb wird im Folgenden der Gegensatz von Kompetenz und Intelligenz thematisiert:

Intelligenz ist eine generalisierbare Eigenschaft, die eine Person unabhängig von Vorwissen und Kontextualisierung dazu qualifiziert, neue Situationen und Problemstellungen erfolgreich zu lösen (Hartig & Klieme, 2006). Für die Intelligenz kann im Gegensatz zur Kompetenz kein Nullpunkt definiert werden, da jedes Individuum kontextunabhängige kognitive Grundfähigkeiten besitzt. Unterschiede in der Intelligenz zwischen den Individuen basieren auf unterschiedlichen kognitiven und biologischen Basiskonzepten (Hartig & Klieme, 2006). Kompetenz und Intelligenz sind Konstrukte, die eng miteinander verknüpft sind und die sich auch inhaltlich überschneiden. Zusätzlich zur Kontextualisierung führen Hartig und Klieme (2006) weitere Kriterien zur Abgrenzung der Begriffe Intelligenz und Kompetenz auf (siehe Tab. 1).

Tab. 1 Gegenüberstellung konzeptueller Charakteristika des Kompetenz- und Intelligenzbegriffs (Hartig & Klieme, 2006)

Kompetenz	Intelligenz
kontextabhängig Fähigkeiten, spezifische Situation und Anforderungen zu bewältigen	generalisierbar Fähigkeit, neue Probleme zu lösen
lernbar wird durch Erfahrung mit spezifischen Anforderungen und Situationen erworben	zeitlich stabil zu bedeutsamen Teilen durch biologische Faktoren determiniert
Leistungsunterschiede bedingt durch Anforderung und Situation	Leistungsunterschiede bedingt durch kognitive Prozesse

In den bundeseinheitlichen Bildungsstandards (Kultusministerkonferenz, 2005a, 2005b, 2005c) sind Kompetenzen domänenspezifisch an die jeweiligen Fächer gebunden und auf bestimmte Kontexte, Situationen und Anforderungen fokussiert. Dadurch ist die Abgrenzung zur Intelligenz möglich (Hartig & Klieme, 2006). Weinert (2001a) schließt die motivationale Orientierung in seiner Definition von Kompetenz ein, plädiert jedoch dafür, kognitive Kompetenzen und motivationale Faktoren getrennt zu erfassen.

3.1.2 Experimentelle Kompetenz

Wissensarten und Denkprozesse bilden die Grundlage für experimentelle Kompetenz. Unterschieden wird dabei deklaratives Wissen, d.h. das experimentbezogene Wissen über inhaltliche sowie methodische Konzepte, und prozedurales Wissen, also die Ausführung eines Experimentes, wie z.B. die Umsetzung der Variablenkontrollstrategie (Gut & Mayer, 2018). Es gibt verschiedene nebeneinander existierende Teilbereiche, um die experimentellen Kompetenzen zu modellieren. Diese sind: wissenschaftliches Argumentieren (Scientific Reasoning), Fähigkeiten und Fertigkeiten (Skills) sowie Prozessfähigkeiten (Process Skills), Problemlösefähigkeiten (Problem Solving) und das Verstehen von Konzepten (Conceptual Understanding) (siehe auch (Gut & Mayer, 2018)).

In der Literatur wird der Begriff Experimentieren nicht immer gleichbedeutend verwendet. Eine allgemein gültige Definition scheint also nicht möglich (Emden, 2011). Experimentieren wird jedoch im deutschen Sprachraum oftmals eng definiert, nämlich als die gezielte handelnde Auseinandersetzung im Interesse eines Erkenntnisgewinns über die Natur mit dem Zweck, Daten zu gewinnen, um diese dann vor dem Hintergrund von Modellen und Theorien zu interpretieren und dadurch Wissen und Erkenntnisse über die Natur abzuleiten (Gut & Mayer, 2018) (siehe Kapitel 3.1.2.3). Dabei wird zur Gliederung des Experimentierprozesses in der Literatur eine Vielzahl möglicher Vorschläge gemacht. Zusammenstellungen finden sich unter anderem bei Emden (2011). Gemeinsam haben alle Vorschläge eine Ideenfindungs-/Hypothesenfindungs-, Durchführung- und eine Auswertungsphase.

In den empirischen Wissenschaften dienen Experimente der Erkenntnisgewinnung. Das Experimentieren wird in diesem Kontext als hypothetisch-deduktive Erkenntnisgewinnung angesehen (Gut & Mayer, 2018). Dabei können aber neben der handelnden Auseinandersetzung mit der Natur auch Methoden zum Experimentieren gezählt werden, die keinen Eingriff in die Natur mit sich ziehen: in der Biologie bspw. das Beobachten und Vergleichen gemäß Wellnitz und Mayer (2016), verfahrensbasierte Tests (Emden & Sumfleth, 2012) in der Chemie oder die technisch-konstruktive Herstellung einer Messvorrichtung (Schreiber et al., 2014) in der Physik.

Die experimentelle Kompetenz wird in der Forschung in zwei unterschiedlichen Ansätzen betrachtet und modelliert (Gut & Mayer, 2018): Beim ersten Ansatz, dem Teilprozessansatz, erfolgt die Erkenntnisgewinnung durch das Abarbeiten von Teilpro-

zessen. Dieser Ansatz beruht auf der Idee, dass der naturwissenschaftliche Erkenntnisgewinnungsprozess im Sinn des *wissenschaftlichen Denkens* in Teilprozesse zerlegt werden kann (Emden, 2011). Die experimentelle Kompetenz umfasst im Sinne des Teilprozessansatzes alle Teilprozesse der experimentellen Arbeitsweise, die zur naturwissenschaftliche Erkenntnisgewinnung führen (Gut & Mayer, 2018). Diese sind:

- a) Fragestellung entwickeln
- b) Hypothesen generieren
- c) Planung und Durchführung des Experiments
- d) Messen, Beobachten und Protokollieren
- e) Daten analysieren (und deren Verallgemeinerbarkeit diskutieren)
- f) Sachgerechte Schlüsse ziehen und diese zu kommunizieren
- g) Reflektieren (Gut & Mayer, 2018).

Die Erkenntnisgewinnung erfolgt durch die adäquate und korrekten Ausführung der einzelnen Teilprozesse (Gut & Mayer, 2018). Das Konstrukt schließt außerdem alle Fähigkeiten und Fertigkeiten ein, die es ermöglichen, diese Teilprozesse korrekt auszuführen, also beispielweise die Kompetenz des Beobachtens oder Vergleichens. Das Ziel des Experimentierens ist die Gewinnung oder Bestätigung von Hypothesen über kausale Zusammenhänge (Wirtz & Schulz, 2012). Die kognitiven Prozesse, auf denen die experimentelle Kompetenz basiert, kann als *wissenschaftliches Denken* (siehe auch Kapitel 3.2.1) betrachtet werden (Gut & Mayer, 2018).

Der Problemlöseansatz ist der zweite Modellierungsansatz zur experimentellen Kompetenz. Die theoretische Rahmung, welche die Grundlage für den Problemlöseansatz legt, beschreibt die experimentelle Kompetenz als die Fähigkeit, eine Vielzahl an Problemstellungen, die sich im Rahmen eines Experiments ergeben, isoliert oder in Kombination, lösen zu können (Gut & Mayer, 2018). Dazu gehören Probleme wie die Untersuchung von kausalen Zusammenhängen mittels Variablenkontrolle, aber auch das Vergleichen von Objekten oder das Testen von Eigenschaften (Gut, 2012; Ruiz-Primo & Shavelson, 1996). Die Kompetenzbeschreibungen beim Problemtypansatz konzentrieren sich auf die Arten des zu lösenden Problems (Gut & Mayer, 2018). Charakteristisch ist dabei die Art der experimentellen Fragestellung, die mit Hilfe der Manipulation von Experimentiermaterial neue, nicht vorhersehbare Daten generiert und zusammen mit dem Vorwissen zur Lösung des Problems beiträgt.

Experimentieren steht als *wissenschaftliches Denken* im Vordergrund. Die Problemtypen umfassen, anders als bei der Erkenntnisgewinnung, nicht nur induktive und deduktive Erkenntnisgewinnungsprozesse, sondern auch die konstruktiven und konstruktiv-technischen Problemstellungen, wie das Entwickeln einer Experimentieranordnung (Gut & Mayer, 2018). Forschende setzen sich bei der Durchführung eines naturwissenschaftlichen Experiments zur Gewinnung oder Bestätigung von Hypothesen mit verschiedenen Problemstellungen auseinander. Bspw. muss die Experimentier- und Messanordnung auf Basis der zu messenden Größen bestimmt und entwickelt werden und müssen während des Experiments Beobachtungen und Daten interpretiert werden (Gut & Mayer, 2018). Unabhängig von der Fragestellung des Experiments ergeben sich isolierte wiederkehrende Problemstellungen wie bspw. die Entwicklung und die Optimierung von Test- und Vergleichsprozedere (Gut & Mayer, 2018). Dabei greifen Forschende auf standardisierte naturwissenschaftliche Methoden zurück, z.B. die Messwiederholung (Gut & Mayer, 2018).

Der Problemtypenansatz, dem eben dieser Problemlöseansatz zu Grunde liegt, beschreibt die experimentelle Kompetenz als Fähigkeit, eine Vielzahl von experimentellen Problemen isoliert oder in Kombination lösen zu können (Gut & Mayer, 2018). Bei diesem Ansatz werden durch Manipulation von Experimentiermaterial in der realen oder virtuellen Umgebung nicht vorhersehbare, neue Daten erzeugt, die zusammen mit dem Vorwissen und den bekannten Informationen zur Lösungen des gestellten Problems beitragen (Gut & Mayer, 2018). Die Lösung der experimentellen Problemtypen (Gut & Mayer, 2018) hat einen Wissenszuwachs über die Natur zum Ziel. Es handelt sich dabei nicht um eine alternative Beschreibung der Erkenntnisgewinnungsmethoden (Wellnitz & Mayer, 2013, 2016) Beobachten, Vergleichen und Experimentieren (siehe Kapitel 3.2.2). Bei der Erkenntnisgewinnung werden die prinzipiellen Prozesse und Zugänge zur Erkenntnis unterschieden. Bei den Problemtypen hingegen wird eine isolierte typische Aufgabenstellung modelliert, die sich durch die praktische Umsetzung eines Experiments und die Anwendung von Methoden in Sinn von „concept of evidence“ anbietet (Gut & Mayer, 2018). Problemtypen umfassen nicht nur induktive und deduktive Erkenntnisgewinnungsprozesse, sondern auch konstruktive und konstruktiv-technische Problemstellungen wie die Entwicklung einer Versuchsanordnung (Gut & Mayer, 2018).

Wissenschaftliches Skizzieren kann die Erkenntnisgewinnungsprozesse, während dem praktischen Handeln und Experimentieren unterstützen. Deshalb wird nachfolgend Erkenntnisgewinnung (siehe Kapitel 3.2) definiert und die verschiedenen Methoden der Erkenntnisgewinnung (siehe Kapitel 3.2.2) beschrieben.

3.2 Erkenntnisgewinnung

Dem Wissen, wie Erkenntnisse gewonnen werden, und dem Können, durch Handlung Erkenntnisse zu gewinnen, wird im naturwissenschaftlichen Unterricht ein zentralerer Stellenwert zugeschrieben. Die naturwissenschaftliche Erkenntnisgewinnung wird ausgehend von Popper (1994) oftmals als hypothetisch-deduktive Vorgehensweise modelliert: Dabei basiert die Erkenntnisgewinnung auf der Verwerfung von Theorien durch empirische Falsifikation. Naturwissenschaftliche Erkenntnisgewinnung wird dabei als hypothetisch-deduktiver Prozess verstanden, der lerntheoretisch als komplexer Problemlöseprozess aufgefasst wird, der von Personen und Situationsvariablen beeinflusst wird (Klahr, 2000; Mayer, 2007). Unter Problemlösen (Klieme et al., 2001) wird das Bestreben verstanden, von einem gegebenen Zustand in einen anderen gewünschten Zustand zu gelangen, wobei eine Barriere zwischen den beiden Zuständen überwunden werden muss (Hussy, 1984). Die Lösung des Problems erfolgt durch verschiedene Prozeduren oder Operationen wie kausales Denken oder Deduktion und Induktion, wodurch diese Barriere überwunden werden kann (Mayer, 2007). Der Problemlöseprozess folgt einer systematischen Abfolge von Prozeduren und kann zu einer Erkenntnisgewinnung führen.

Nach Bybee (2002) ist diese naturwissenschaftliche Erkenntnisgewinnung ein Kernstück der naturwissenschaftlichen Bildung. Daher wurde sie in den deutschen Curricula (Kultusministerkonferenz, 2005b) als Denk- und Arbeitsweise einbezogen. Der Kompetenzbereich Erkenntnisgewinnung in den deutschen Curricula ist ein wesentlicher Teil des Konstrukts „Science“, welches aus drei Elemente besteht: Kenntnisse naturwissenschaftlicher Inhalte (kann durch den Kompetenzbereich Fachwissen abgedeckt werden), Verständnis und Durchführung des naturwissenschaftlichen Erkenntnisprozesses sowie Wissenschaftsverständnis bezüglich der Naturwissenschaften (Köller et al., 2008).

Im Rahmen des Projekts zur Evaluation der Standards (ESNaS) wurde der Kompetenzbereich Erkenntnisgewinnung in drei Teilbereiche unterteilt (Kremer et al.,

2012a): Naturwissenschaftliche Untersuchungen, naturwissenschaftliche Modellbildung und wissenschaftstheoretische Reflexion. Diese drei Teilbereiche enthalten jeweils die drei oben genannten Elemente des Konstrukts „Science“, d.h. Kenntnisse naturwissenschaftlicher Inhalte, Verständnis und Durchführung des naturwissenschaftlichen Erkenntnisprozesses und Wissenschaftsverständnis bezüglich der Naturwissenschaften.

Im Kompetenzbereich Erkenntnisgewinnung für das Fach Chemie wurden in Deutschland für den Mittleren Schulabschluss acht Standards mit ebenfalls prozessorientierter Ausrichtung ausformuliert (Kultusministerkonferenz, 2005b), welche mit „Erkenntnisse können gewonnen werden, indem experimentelle und andere Untersuchungsmethoden sowie Modelle nutzen werden“ (Kultusministerkonferenz, 2005b) zusammengefasst werden können. Dies geschah jedoch ohne Spezifizierung der Untersuchungsmethoden. Die Aufgabe der fachdidaktischen Forschung war es, die Kompetenzen zu konkretisieren und dadurch diagnostizierbar zu machen (Nehring et al., 2016).

Im Folgenden werden Erkenntnisgewinnung als naturwissenschaftliche Arbeitsweisen und wissenschaftliches Denken miteinander verknüpft und auf diese Weise die Kompetenzen konkretisiert.

3.2.1 Naturwissenschaftliche Arbeitsweisen und wissenschaftliches Denken

Nehring et al. (2016) verknüpfen matrixartig die *naturwissenschaftlichen Arbeitsweisen* „Modellieren“, „Experimentieren“ und „Beobachten, Vergleichen, Ordnen“ mit dem wissenschaftlichen Denken, der adäquaten Durchführung der einzelnen Teilprozesse des Teilprozessansatzes (siehe Tab. 2). Das Kompetenzmodell zur Vernetzung der Erkenntnisgewinnung (VerE-Modell) (Nehring et al., 2016) definiert Teilkompetenzen auf Grundlage der Kombination der beiden Strukturdimensionen wissenschaftliches Denken und naturwissenschaftliche Arbeitsweisen (Nehring, 2014; Nehring et al., 2016; Nowak et al., 2013). Das wissenschaftliche Denken umfasst naturwissenschaftliche Untersuchungen während die Dimension naturwissenschaftliche Arbeitsweisen verschiedene Arten von Untersuchen definiert, die entsprechend dieser Schritte umgesetzt werden und der Erkenntnisgewinnung dienen (Nehring et al., 2016). Zum wissenschaftlichen Denken zählen in Anlehnung an das SDDS-Modell von Klahr (2000) die Bildung von Fragenstellung und Hypothese, die Planung und

Durchführung von Untersuchungen und deren Auswertung und Reflexion (siehe Tab. 2). Es handelt sich dabei um die Teilprozesse experimenteller Handlung (Gut & Mayer, 2018), durch deren korrekte Ausführung Erkenntnis gewonnen werden kann. Die naturwissenschaftlichen Arbeitsweisen beschreiben die unterschiedliche Umsetzung von naturwissenschaftlichen Untersuchungen, nämlich Beobachten, Vergleichen, Ordnen, Experimentieren und Nutzen von Modellen (Nehring et al., 2016) (siehe Tab. 2). Durch die adäquate Durchführung der *naturwissenschaftlichen Arbeitsweisen* kann ebenfalls Erkenntnis gewonnen werden.

Tab. 2 Struktur der Erkenntnisgewinnung in der Chemie nach Nehring et al. (2016)

<i>Naturwissenschaftliche Arbeitsweisen</i>			
Modelle nutzen			
Experimentieren			
Beobachten, Vergleichen, Ordnen			
	Fragestellung & Hypothesen	Planung & Durchführung	<i>Wissenschaftliches Denken</i>

Die verschiedenen Methoden der Erkenntnisgewinnung entsprechen den unterschiedlichen fachtypische Arbeitsweisen in den Naturwissenschaften. Die Arbeitsweisen Experimentieren und Beobachten beispielsweise unterscheiden sich dadurch, dass andere Erkenntnisse gewonnen werden: Beim Experimentieren ist es der Aufschluss über kausalen Zusammenhänge (Gut & Mayer, 2018), und beim Beobachten sind es Kenntnisse über die nominale oder ordinale Einordnung von Merkmalsausprägungen (Gut & Mayer, 2018).

Gemeinsame Kriterien der Methoden der Erkenntnisgewinnung (siehe Kapitel 3.2.2) und der Problemtypen führen zu einer adäquaten Abfolge von Teilprozessen (siehe Kapitel 3.1.2). Die Herstellung von sinn- und sachlogischen Zusammenhängen, wie bspw. die Abfolge von Teilprozessen, wird den wissenschaftlichen Denkweisen zugeordnet. Nachfolgend werden die Methoden der Erkenntnisgewinnung im Sinne der naturwissenschaftlichen Arbeitsweisen näher beleuchtet.

3.2.2 Methoden der Erkenntnisgewinnung

Die Erkenntnisgewinnung umfasst die Untersuchungsmethoden des Experimentierens (Variablenkontrollstrategie) wie auch das praktische Handeln des Beobachtens

und des Vergleichs (Kultusministerkonferenz, 2005b). Die Biologie nutzt das Beobachten, das Vergleichen und das Experimentieren als grundlegende Methoden der Erkenntnisgewinnung (Wellnitz & Mayer, 2013). Beim Beobachten, Vergleichen und Ordnen hingegen steht der theorie- bzw. kriteriengeleitete Umgang mit Objektmerkmalen im Vordergrund (Nehring et al., 2016) (siehe Kapitel 3.2.2.1. und 3.2.2.2). Unter dem Begriff Experimentieren wird die Anwendung der Variablenkontrollstrategie und damit die gezielte Manipulation von Variablen verstanden (Nehring et al., 2016). Untersucht wird eine planmäßige und systematische Variation einer unabhängigen Variablen, wobei der Einfluss auf die abhängige Variable beobachtet wird (Nehring et al., 2016) (siehe Kapitel 3.2.2.3).

Ebenso gehört das Modellieren und die Nutzung von Modellen zu den Erkenntnisgewinnungsmethoden. Bei der Anwendung von Modellen mit dem Zweck der Erkenntnisgewinnung werden diese genutzt, um Vorhersagen über Eigenschaften bzw. Zusammenhänge über das Original/Phänomen zu machen (Nehring et al., 2016). Das Modell wird durch die Überprüfung der Hypothesen anhand aus der Natur gewonnener Daten vorgenommen und kann zu einer Änderung des Modells führen. Erkenntnisgewinnung durch Modellieren oder durch Nutzung von Modellen kann eine experimentelle Handlung wie Modellexperimente einschließen. Modelle werden auch genutzt, wenn man im Rahmen hypothetisch-deduktiven Experimentierens eine Hypothese zu einem Zusammenhang aufstellt und begründet. Die Skizze einer Planung oder einer experimentellen Handlung, wie bspw. eine Schnittzeichnung, oder die Skizze eines Stromkreislaufs mit einer Schaltskizze können jeweils sehr idealisiert sein und als Modellzeichnungen betrachtet werden. Skizzen von der Beobachtung eines Phänomens gleichen eher Fotografien, welche die Realität wiedergeben. Im Forschungsfeld wird das Modellieren mit praktischem Handeln und Experimentieren verknüpft, die Nutzung von Modellen zu den Erkenntnisgewinnungsmethoden wie auch Reflexionen werden jedoch im naturwissenschaftlichen Unterricht kaum dokumentiert, wohingegen die Dokumentation von experimentellen Handlungen sehr häufig ist. Daher wurde das *wissenschaftliche Skizzieren* im Rahmen dieser Dissertation wie folgt eingeschränkt und bewusst auf die Modellierungsaspekte bei den Erkenntnisgewinnungsprozessen verzichtet: *Wissenschaftliches Skizzieren* bezeichnet die Fähigkeiten und Fertigkeiten, während des Experimentierprozesses objektive Skizzen anzufertigen, die als förderliche Basis der Erkenntnisgewinnung dienen.

Im Folgenden werden die Methoden der Erkenntnisgewinnung: Beobachten, Vergleichen und Experimentieren beleuchtet, da durch diese Methoden in experimentellen Handlungen Erkenntnisse gewonnen werden können, diese aber auch im naturwissenschaftlichen Unterricht als Dokumentier-Anlässe genutzt werden. Diese Fokussierung deckt sich auch mit der von Nehring et al. (2016) gemachten Verknüpfung von naturwissenschaftlichen Arbeitsweisen mit dem wissenschaftlichen Denken.

3.2.2.1 *Beobachten*

Alltagsbeobachtungen finden eher zufällig statt und unterliegen keinem genaueren Plan, sie finden spontan und unter subjektiven Aspekten statt (Bortz & Döring, 2007). Heath (2018) beschreibt das Beobachten, verbunden mit Skizzieren, als einen Prozess des konzentrierten Sehens, als eine intensive und eindringliche Form der Untersuchung, die ein Bild sowohl auf der Netzhaut in der Erinnerung als auch auf dem Papier fixiert, und zwar auf eine Weise, die sich deutlich von der Aufnahme eines Fotos oder einer schriftlichen Notiz unterscheidet. Im wissenschaftlichen Sinn wird unter Beobachten eine systematische Vorgehensweise verstanden, die geplant und anhand von Kriterien stattfindet (Bortz & Döring, 2007). Dabei erfordert das Sehen kein Sehen mit dem Auge, denn auch in der Astronomie, Kosmologie oder der Atom- und Teilchenphysik wird beobachtet (z.B. mit einem Elektronentunnelmikroskop). Beim wissenschaftlichen Beobachten versucht man die Natur zu erkennen, ohne manipulativ einzugreifen.

Eine Unterscheidung zwischen Betrachten und Beobachten wird vornehmlich durch die Biologiedidaktik vertreten (Gropengießer et al., 2010): Während beim Betrachten die Auseinandersetzung mit unbewegten und unveränderlichen Objekten im Untersuchungsmittelpunkt steht, werden beim Beobachten dynamische Vorgänge und Veränderungen fokussiert. Mittels Beobachtungen lassen sich Vorgänge und Bewegungen innerhalb einer bestimmten Zeitspanne bewusst erfassen und beschreiben. Bei Originalbegegnungen lässt sich erfahrungsbasiertes Wissen aufbauen (Hild et al., 2015).

Ziel der naturwissenschaftlichen Methode Beobachten ist die nominale oder ordinale Einordnung von Merkmalsausprägungen (Gut & Mayer, 2018). Problemstellung des Beobachtens ist es, Phänomene anhand gegebener Kategorien zu beschreiben und zu vergleichen. Zentral für Beobachtungen ist auch, sicherzustellen, dass mehrere Beob-

achter unter den gleichen Bedingungen zu demselben Resultat gelangen (Sommer et al., 2018).

Beobachten ist eine komplexe, vielschichtige Tätigkeit (Oguz & Yurumezoglu, 2007) die sich in Teilkompetenzen aufspalten lässt. Unter Berücksichtigung der generellen wissenschaftspropädeutischen Teilkompetenzen (Harms et al., 2004) sind konvergentes und divergentes Denken, Verbalisieren eines Sachverhaltes und das naturwissenschaftliche Argumentieren für die Teilkompetenzen des systematischen Beobachtens relevant (Kohlhauf, 2012).

Die Teilkompetenz systematische Beobachtung kann einerseits Teil einer Untersuchung sein, andererseits auch eine eigene wissenschaftliche Methode mit einem spezifischen Beobachtungsdesign darstellen. Sind systematische Beobachtungen Teil einer Untersuchung, dienen sie der objektiven Beschreibung von experimentell gewonnenen Daten (Wellnitz & Mayer, 2016). Beobachten gilt als Grundlage für andere naturwissenschaftliche Erkenntnismethoden, aber auch als eigenständige Forschungsmethode, die besonders in der Biologie zentral ist (Harms et al., 2004). Gut und Mayer (2018) bezeichnen dies als Teilprozess des Experimentierprozesses im Sinne des Teilprozessansatzes. Des Weiteren können Beobachtungen als gezielte Methode der Datenerhebung zur Beantwortung einer Fragestellung und dadurch zur Prüfung von Hypothesen genutzt werden (Bortz & Döring, 2007). Im Problemtypansatz von Gut und Mayer (2018) ist Beobachten ein eigener Problemtyp, der in der Biologie häufig angewandt wird.

Die Beobachtungskompetenz gemäß Kompetenzmodell von Kohlhauf, Rutke und Neuhaus (Kohlhauf et al., 2010) unterscheidet inzidentelles Beobachten, unsystematisches Beobachten und systematisches Beobachten. Systematische Beobachtungen stellen in der Regel den ersten Zugang zu einem Phänomen im Rahmen der naturwissenschaftlichen Erkenntnisgewinnung dar. Beobachten kann durch Präkonzepte und Vorwissen beeinflusst werden (Kohlhauf, 2012). Inzidentelle und unsystematische Beobachtungen sind Vorstufen zum Erreichen einer ausgeprägten Beobachtungskompetenz. Sind Unterschiede und Gemeinsamkeiten zweier Phänomene zu identifizieren stellen Erweiterungen zum korrekten und vollständigen Beobachten eines Phänomens dar (Gott & Welford, 1987). Die Beschreibung und Dokumentation von Beobachtetem ist eine bedeutende Kompetenz des wissenschaftlichen Beobachtens (Wellnitz & Mayer, 2008).

Lernende machen wenig Beobachtungsfehler, jedoch viele Fehler, was die Menge der möglichen Beobachtungskategorien angeht (Gott & Welford, 1987). Sie schreiben bspw. viel über Grösse, Farbe etc. eines Vogels, aber nichts über die Form des Vogels. Dies geschieht, weil sie von ihren Präkonzepten beeinflusst werden und zudem nicht wissen, worauf sie ihren Fokus richten müssen (Gott & Welford, 1987). Außerdem geschehen viele verschiedene Vorgänge (beispielweise Farbwechsel und Gasentwicklung) oftmals simultan. Durch kategoriengeleitete Aufträge kann der Fokus des Betrachters auf das Relevante gerichtet werden. So wird ersichtlich, worauf die Beobachtenden den Fokus legen sollen. Durch Leitfragen wie: „Was kannst du zur Farbe im Messbecher sagen? Beschreibe und skizziere, was du beobachtest“ erhalten die Lernenden eine deutliche Anleitung, mit welchem Fokus sie beobachten sollen (Hild et al., 2015). Beobachtungen sollten zielorientiert, fokussiert und kategorienorientiert sein (Hild et al., 2015; Wellnitz & Mayer, 2016).

Die Beobachtungskompetenz steigt mit zunehmendem Bildungsstand signifikant an (Kohlhauf, 2012). Jedoch ignorieren Schülerinnen und Schüler der 9. Jahrgangsstufe noch immer das Fachwissen zu einem gegebenen Problem zu zwar bei der Methode des Beobachten, beim Vergleichen und Experimentieren und beim Formulieren der Fragestellung oder der Hypothese (Arnold et al., 2010). Sogar angehende Lehrpersonen führen Beobachtungen nicht immer systematisch durch oder halten wissenschaftlichen Kriterien nicht ein (Oguz & Yurumezoglu, 2007). Die Beobachtungen werden ähnlich wie Alltagsbeobachtungen durchgeführt. Weder Forschungsfragen noch zu messende Merkmale werden festgelegt (Oguz & Yurumezoglu, 2007).

Das Skizzieren leistet einen wichtigen Beitrag zur Verbesserung des Beobachtungsprozesses, weil Vorgänge für die Darstellung präzise beobachtet werden müssen. Systematisches Beobachten und damit die Beobachtungsfähigkeit ist die Grundlage für die Erkenntnisgewinnungsprozesse während des Skizzierens. Skizzieren bedingt also genaues Beobachten. Eine nicht ausgeprägte Beobachtungskompetenz kann Auswirkungen auf die Kompetenz des Skizzierens haben, da die Lernenden das, was sie nicht beobachtet haben, auch nicht bildlich darstellen werden. Beobachten kann durch Präkonzepte und Vorwissen beeinflusst werden, somit ist es abhängig von der Wahrnehmung. Durch einen gerichteten Fokus, wie mithilfe von Beobachtungskriterien, kann die Objektivität der Beobachtung und dadurch der Skizze erreicht werden.

3.2.2.2 *Vergleichen*

Beim Vergleichen als Methode der Erkenntnisgewinnung werden Systeme oder Phänomene auf Gemeinsamkeiten und Unterschiede überprüft (Wellnitz & Mayer, 2016). Dabei handelt es sich um effektbasiertes Vergleichen. Beim kriteriengeleiteten Vergleichen hingegen werden mindestens zwei Vergleichsobjekte über ein Vergleichskriterium zueinander in Beziehung gestellt (Gott & Welford, 1987; Hammann, 2002) und bezüglich Ähnlichkeiten oder Gemeinsamkeiten und Unterschiede erfasst (Brezmann, 2004). Ein Vergleich besitzt folgende Vergleichsmomente:

- a) einen Anstoß zum Vergleich (Frage, Problem oder Hypothese, die den Vergleich auslöst)
- b) ein Vergleichskriterium
- c) Vergleichsobjekte
- d) ein Vergleichsergebnis (Ottini, 2018)

Verglichen werden ausgewählte Kriterien. Ob Objekte oder Vorgänge ähnlich sind, hängt davon ab, welche Eigenschaften sie besitzen, aber auch davon, wer den Vergleich anstellt und zu welchem Zweck. Ähnlichkeiten besitzen immer auch eine Kontextabhängigkeit (Hammann, 2002). Objekte können anhand einer gegebenen Eigenschaft experimentell (ohne direkte Messung der Eigenschaft) verglichen werden. Die Lernenden sollen die Einsicht gewinnen, dass Gesetzmäßigkeiten sich nicht an Einzelobjekten gewinnen lassen (Hild et al., 2015). Das Vergleichskriterium ist entscheidend für das Vergleichsergebnis. Dieses muss deshalb in der Fragestellung, die den Fokus auf Unterschiede und Gemeinsamkeiten hin lenkt, genannt werden (Gott & Welford, 1987).

Analog zum systematischen Beobachten ist der kriteriengeleitete Vergleich Grundlage für die Erkenntnisgewinnungsprozesse aus Vergleichen während des Skizzierens. Skizzieren bedingt also sowohl genaues Beobachten als auch die Methodenkompetenz des kriteriengeleiteten Vergleichens.

Es bestehen also Abhängigkeiten der Teilkompetenzen, wie bspw. der Beobachtungs- und Vergleichskompetenz, zu den (Teil-)Kompetenzen des Skizzierens. Nach Alfieri et al. (2013) gilt das Vergleichen grundsätzlich als einfacher als das Beobachten.

3.2.2.3 *Experimentieren*

Das Experimentieren wurde bereits im Kapitel 3.1.2 im Rahmen der experimentellen Kompetenz besprochen. Beim Experimentieren werden Zusammenhänge zwischen gegebenen Variablen untersucht. Die Erkenntnis, dass Zusammenhänge als Abhängigkeit einer abhängigen Variablen von unabhängigen Variablen betrachtet werden kann, ist Voraussetzung für die Lösung des Problems des Experimentierens. Das benötigte Methodenwissen ist die Variablenkontrolle, bei der als Strategie einfache Abhängigkeiten isoliert und untersucht werden, woraus Erkenntnisse über die qualitative Ausprägung von Merkmalszusammenhängen gewonnen werden (Gut & Mayer, 2018). Das Verständnis von kontrollierten Experimenten wie auch das Wissen über kontrollierte Experimente ist bei der Planung, Durchführung und Auswertung von Experimenten zentral. Bei kontrollierten Experimenten wird die Variablenkontrollstrategie (VKS) eingesetzt. Während eine Variable verändert wird, bleiben alle anderen unverändert (Chen & Klahr, 1999). Bei der experimentellen Erkenntnisgewinnung gilt die VKS als ein zentraler Aspekt (Schwichow et al., 2016).

Koslowski (1996) untersuchte in allgemeinen wie auch naturwissenschaftlichen Situationen das wissenschaftliche Argumentieren (Scientific Reasoning) von Lernenden über Evidenz und Schlussfolgerungen. Sie interessierte sich dafür, wie Lernende mit Evidenzen umgehen, die eine gegebene Erklärung widerlegen, also wie Lernende mit dem Prozess der Hypothesenwiderlegung umgehen. In der Argumentation über Evidenz und Erklärung berücksichtigten Lernende die Rolle der Kovarianz und der Nicht-Kovarianz und ebenso die möglichen Mechanismen von unabhängigen und abhängigen Variablen nicht adäquat (Koslowski, 1996). Koslowski (1996) untersuchte in einer Serie von Experimenten, wie Lernende durch zusätzliche Informationen in ihren Evidenzen beeinflusst werden. Obwohl der Diskurs über Unterschiede in sozial- und naturwissenschaftlichen Situationen nicht geführt wurde, stellt Koslowski den Einfluss von abhängigen und unabhängigen Variablen als eine Basis der VKS dar.

Um den Einfluss des Fachwissens auf die Ausprägung der VKS beurteilen zu können, untersuchten Schwichow und Nehring (2018) mittels Strukturgleichungsmodellierung die Datensätze zweier unabhängiger Querschnittstudien der Jahrgangsstufe 5 bis 13 (Physik) bzw. 8 bis 12 (Chemie/ Biologie) im Gymnasium. Sie konnten eine mittlere signifikante Korrelation zwischen der Jahrgangsstufe und dem Fachwissen wie auch zwischen der VKS und dem Fachwissen für alle drei Fachkontexte zeigen. Für

alle drei Fachkontexte (Biologie, Chemie und Physik) ging eine höhere Kompetenzausprägung mit höherem Fachwissen und höherer Jahrgangsstufe einher. Dies bestätigte den erwarteten positiven Zusammenhang zwischen Fachwissen und VKS beim Experimentieren (Schwichow & Nehring, 2018). Dreifaktorielle Modelle zeigten keinen signifikanten Zusammenhang (bzw. nur einen schwach signifikanten für das Fach Chemie) zwischen Jahrgangsstufe und VKS. Das Fachwissen wirkt signifikant auf die VKS, und die Jahrgangsstufe wirkt signifikant auf das Fachwissen. Der Einfluss des Fachwissens aus einem Fach auf die VKS eines anderen Fachs zeigte keinen direkten signifikanten Pfad, das Fachwissen der Biologie wirkte jedoch stärker auf die VKS der Chemie als das chemische Fachwissen. Eine mögliche Erklärung sahen Schwichow und Nehring (2018) in einer indirekten kausalen Abhängigkeit zwischen Fachwissen und VKS. Ursachen dieser Korrelation könnten allgemeine kognitive Fähigkeiten sein, welche die Beantwortung von Fachwissens- und VKS-Aufgaben positiv beeinflussen. Mit dem vorliegenden Datenmaterial des KFT (Heller & Perleth, 2000) konnte diese Vermutung jedoch nicht gestützt werden, da kein direkter Zusammenhang zwischen den kognitiven Fähigkeiten und der VKS-Fähigkeit vorlag. Zur Erklärung wurden Perspektiven aufgestellt, darunter die Perspektive, dass Fachwissen und VKS nicht trennbar seien, was für eine allgemeine naturwissenschaftliche Kompetenz spräche. Diese Perspektive könnte auch durch die von Klos et al. (2008) bestätigte mittlere Korrelation zwischen Fachwissen und experimenteller Kompetenz für die 12. Jahrgangsstufe gestützt werden.

Es gilt als unklar, inwiefern die Kompetenzausprägungen älterer Lernender auf zunehmendes Fachwissen oder zunehmende prozessorientierte Kompetenzausprägung zurückzuführen sind (Schwichow & Nehring, 2018). Es konnten nämlich weder genaue Informationen zum Wissenserwerb und zu genuinem Kompetenzerwerb im Laufe der Schulkarriere noch zur Wirkung des naturwissenschaftlichen Unterrichts gemacht werden (Nehring & Schwichow, 2020). Nehring und Schwichow (2020) untersuchten, inwiefern prozessorientierte Kompetenzen (Methodenkompetenz wie die VKS) und Fachwissen als inhaltsorientierte Kompetenzen (VKS und Fachwissen sind unabhängig) oder als getrennte Konstrukte angesehen werden können; dies in einer Re-Analyse eines Datensatzes einer Querschnittsanalyse zur Variablenkontrollstrategie. Sie zeigten, dass zwischen den Kompetenzbereichen Fachwissen und prozessorientierten Kompetenzen aus dem Bereich der Erkenntnisgewinnung ein nicht zu vernachlässigender Zusammenhang bestand.

Da in den Jahrgangsstufen 2 bis 4 große Veränderungen der Fähigkeit der VKS zu erwarten sind (Chen & Klahr, 1999), konzentrierten sich Salim et al. (2020) auf diese Altersklassen und konnten zeigen, dass bereits Grundschulkinder durchaus ein echtes Verständnis für die VKS erreichen und dies auch in schriftlichen Testaufgaben wiedergeben können. Ebenfalls mit diesen Klassenstufen beschäftigten sich Brandenburger et al. (2020) und zeigten, dass Lernende der höheren 4. Jahrgangsstufe schwerere Items erfolgreich bearbeiten können, sowie auch, dass knapp 80 % der Lernenden die korrekten Experimente-Paare identifizierten. Neben bekannten Fehlern zeigten nur wenige Lernende (9.2 %) kein eindeutiges Antwortverhalten. Deutlich schlechter schnitten die Lernenden jedoch bei der Begründung ihrer Interpretation ab. Sodian und Meyer (2013) zeigten, dass Kinder bereits mit sechs Jahren zwischen Hypothesentest und Effektproduktion unterscheiden können, über ein implizites Verständnis der Variablenkontrollstrategie (VKS) verfügen und die Entwicklung des expliziten Verständnisses beginnt. Unter implizitem Wissen wird das Wissen verstanden, dass man etwas anwenden, jedoch nicht erklären kann, was man warum tut (Kluwe, 2006). Kinder sind zwar in der Lage, mit implizitem Verständnis über die VKS diese anzuwenden, können aber ihre Entscheidung nicht auf der Grundlage der VKS begründen (Sodian & Mayer, 2013). Mannel et al. (2015) konnten die in der Literatur beschriebenen Schwierigkeiten grade von jüngeren Lernenden insbesondere des mittleren Leistungsbereichs replizieren.

Außerdem besteht eine Altersabhängigkeit für das Verständnis der VKS (Sodian & Mayer, 2013). Während nur ein Drittel der 5. Jahrgangsstufe in der Lage war, ein einfaches Experiment zu kontrollieren, waren es bei den 17-jährigen bereits achtzig Prozent. Je nach Auswahl der kontrollierten Experimente, waren es bei der 3. Jahrgangsstufe dreißig Prozent, bei den 4. und 5. Jahrgangsstufe schon sechzig Prozent und bei den 6. Jahrgangsstufe bereits achtzig Prozent.

Das Experimentieren als Erkenntnisgewinnungsmethode erfordert das Verständnis der VKS, das sich erst mit zunehmendem Alter entwickelt. Ob diese Kompetenzausprägungen jedoch auf zunehmendes Fachwissen oder zunehmende prozessorientierte Kompetenzausprägung zurückzuführen sind und inwiefern die Kompetenzausprägungen von allgemeinen kognitiven Fähigkeiten geprägt sind, ist jedoch unklar. Für das Skizzieren von Experimenten bedeutend ist jedoch die Tatsache, dass Lernende zwar mit dem implizitem Verständnis über die VKS diese zwar anwenden

können, jedoch ihre Entscheidung nicht auf der Grundlage der VKS begründen können (Sodian & Mayer, 2013). Das könnte bedeuten, dass Lernende zwar eine VKS darstellen können, diese jedoch nicht verstehen.

3.3 Repräsentationskompetenz

Multiple External Representation (MER) sind externe Repräsentationen, die aus mehreren Bestandteilen wie Bild, Zeichnung, Grafik oder Text bestehen. Die Repräsentationen stellen Werkzeuge fürs Argumentieren, Lernen und Kommunizieren dar (Nitz et al., 2014). Die Kompetenz im Umgang mit MER wird Representational Competence (Repräsentationskompetenz, RC) genannt. Die RC ist eine umfassende Kompetenz, die alle Kompetenzen im Umgang mit Repräsentationen beschreibt (diSessa, 2002), und daher sind Repräsentationsfähigkeiten eine notwendige Fachkompetenz (Goodwin, 1994). RC ist ein Set von Fähigkeiten, die einer Person den reflektierten Gebrauch einer Vielfalt von Repräsentationen und Visualisierungen erlaubt, einschließlich des Erstellens von (*wissenschaftlichen*) *Skizzen*. Dieses Set an Fähigkeiten wird nachfolgend als strukturierte Aufzählung ausführlich besprochen. Weil mit MER unter anderem neue Modelle entwickelt, Voraussagen gemacht, Schlussfolgerungen gezogen und Übereinstimmungen gebildet werden können, bieten diese eine förderliche Basis zur Erkenntnisgewinnung (Pande & Chandrasekharan, 2017) und sind ein wichtiger Bestandteil der naturwissenschaftlichen Praxis.

Das Wissen in der Chemie kann laut Johnstone (1982) auf drei Levels betrachtet werden, auf der makroskopischen Ebene, die zur Beschreibung der Eigenschaften (Farbe, Dichte, Brennbarkeit etc.) dient, auf der symbolische Ebene (Formeln, Gleichungen, Konventionen etc.) und auf der submikroskopischen Ebene (molekulare Interaktionen). Um Chemie zu erlernen oder zu betreiben, wird oft mit verschiedenen Repräsentationen gearbeitet, beispielsweise mit chemischen Gleichungen, Graphen, Diagramme und Darstellungen von Experimentieraufbauten und Molekülen (Johnstone, 1982; Pande & Chandrasekharan, 2017). Das Erlernen, Unterrichten und Anwenden von Chemie wäre ohne Repräsentationen nicht möglich. Das Integrieren von MER, um chemische Phänomene zu verstehen, MER zu verknüpfen und in andere MER zu überführen oder auch, um neue angemessene MER herzustellen, stellt die wesentlichen RC in der Chemie dar. Damit werden nicht sichtbare Prozesse sichtbar gemacht.

Durch die Anwendung von Inskriptionen, also mit Text und Grafiken ergänzte Skizzen, lernen Studierende ein neues Werkzeug der Naturwissenschaften kennen und entwickeln dadurch eine holistische Sicht auf naturwissenschaftliche Phänomene (Goodwin, 1994). MER und RC werden in der Chemie in unterschiedlichsten Kontexten angewendet, thematisch bspw. zur Darstellung von Konzentrationsverhältnissen oder von Reaktionsmechanismen. Das Periodensystem der Elemente (PSE), chemische Gleichungen, Animationen und Simulationen, molekulare Diagramme und beobachtbare Größen werden häufig als MER dargestellt.

Die Nutzung von naturwissenschaftlichen Repräsentationen ist jedoch stark beeinflusst vom Vorwissen. Dies zeigt sich bspw. darin, dass Lernende mit einem großen Vorwissen über organische Chemie deutlich besser in der Lage sind, Rückschlüsse aus neuen Visualisierungen zu ziehen, als Lernende mit einem geringeren Vorwissen (Hinze et al., 2013). Hill und Sharma (2015) hingegen konnten keinen signifikanten Unterschied bezüglich Vorwissen und Übersetzungen von einer in eine andere Repräsentation zwischen Erst-, Zweit- oder Dritt-Jahr-Physikstudierenden aufzeigen. Auch Chang und Tzeng (2018) zeigten, dass der Besuch von mehr Kursen zum selben Thema keinen Einfluss auf die Performanz bezüglich Repräsentationen hat. Der Faktor Vorwissen in Bezug auf Leistungen in der Repräsentationskompetenz ist somit nicht abschließend geklärt.

Schwierigkeiten bereiten den Lernenden vor allem Visualisierungen, Verknüpfungen zwischen verschiedenen MER, Transformationen von einem MER in ein anderes, insbesondere zwischen statischen und dynamischen Darstellungen, sowie die konzeptionelle Einpassung zwischen verschiedenen MER und die Abstraktion von Repräsentationen (Pande & Chandrasekharan, 2017).

DiSessa und Sherin (2000) führen allgemeine Merkmale der RC auf. Die Taxonomie zur Analyse, was Lernende von Repräsentationen wissen, umfasst folgende Begriffe:

- a) Konstruktives Wissen: Über welche Fähigkeiten und Ideen verfügen die Lernenden, um neue Repräsentationen anzufertigen?
- b) Kritikfähigkeit: Welches Wissen haben die Lernenden, um die Qualität von Repräsentationen zu beurteilen?
- c) Funktionalität: Was wissen Studierende über die Funktion von Repräsentationen?

- d) Lernen: Verfügen Lernende über Strategien, um mittels Repräsentationen zu lernen? (diSessa & Sherin, 2000)

Die Entwicklung und Förderung der Repräsentationskompetenzen erfolgt über Anwendungen, wie z.B. die Anfertigung von Skizzen mit Alltagsbezug etwa die eindimensionale Bewegung eines Objekts, Abfolgen von Sequenzen mit dem Auftrag „Erzähl eine Geschichte in Bildern“ oder die Auseinandersetzung mit Landschaften, die in Kartenform gezeichnet werden (diSessa & Sherin, 2000). Die Lernenden eignen sich dabei die Fähigkeit an, Abfolgen von Bewegungen und Vorgängen dazustellen, was für die Veranschaulichung von chemischen Phänomenen zentral ist. Dies haben Kozma und Russell (1997) als Teilaspekt der RC in der Chemie aufgeführt: Diese Teilkompetenzen der RC sind einerseits das Nachdenken und Handeln in Bezug auf chemische Phänomene, darunterliegende physikalische Zugänge und Prozesse und andererseits das Kommunizieren in Bezug auf die chemischen Phänomene (Gebre & Polman, 2016; Kozma, 2003; Kozma et al., 2000; Kozma & Russell, 1997, 2005; Madden et al., 2011). Kozma und Russell (1997) haben Skills (Fähigkeiten und Fertigkeiten) definiert, über die Studierende im Sinne von Repräsentationskompetenz in der Chemie verfügen sollten:

- a) mit Hilfe von Repräsentationen chemische Phänomene darstellen
- b) adäquate Repräsentation für einen bestimmten Zweck erstellen oder auswählen
- c) Merkmale der einzelnen Repräsentationen finden und analysieren (bspw. den Peak eines Graphen)
- d) eine Repräsentation in eine andere überführen
- e) eine geeignete Repräsentation auswählen oder herstellen
- f) begründen, warum eine bestimmte Repräsentation geeigneter ist als eine andere
- g) beschreiben, wie verschiedene Repräsentationen dasselbe auf verschiedene Arten aussagen
- h) darlegen, dass eine Repräsentation ein Phänomen darstellt, sich aber klar von diesem unterscheidet (Modellkritik)
- i) Repräsentationen für den Diskurs von Aussagen, Schlussfolgerungen und Voraussagen nutzen

Um RC messen zu können, wird ein Kompetenzmodell mit Kategorien benötigt. Chang (2018) entwickelte ein Kodierschema mit vier Kategorien, um allgemeine Repräsentationskompetenzen von Studierenden zu erstellen oder transformierte Reprä-

sentationen zu beurteilen. Diese vier Kategorien wurden unterschiedlich gewichtet. Bei der Kategorie „Einsatz von dynamischen Repräsentationen“ wurde die Anzahl Skizzen erfasst, die ein Student erstellt. Bei der Kategorie „Einsatz von multiplen Repräsentationen“ wurde berücksichtigt, ob die Skizzen alle vier Komponenten, nämlich Text, die symbolische Ebene, die makroskopische Ebene und die Teilchenebene enthalten. Die Kategorie „Einsatz von adäquaten naturwissenschaftlichen Konzepten“ erfasste die Anzahl korrekt verwendeter Komponenten aufgeteilt nach Objekten/Labels, Positionen/Verbindungen und Prozessen. Die letzte Kategorie umfasst die „Nutzung von sichtbar machenden Strategien“ wie Farben, Hervorhebungen und animierter Rahmen (durch das genutzte Programm).

Die Entwicklung der RC bei Jugendlichen führt zu einem besseren Verständnis und einer besseren Kommunikation von naturwissenschaftlichen und mathematischen Ideen (diSessa, 2004) und fördert die innovativen Fähigkeiten und die Motivation zum Erlernen des Fachinhalts (Azevedo, 2000; Van Meter & Garner, 2005). Zudem führt RC zu einer Verbesserung der Zuordnung von inneren (mental) und äußeren Modellen (Zhang, 1997) wie auch zu einer Verbesserung qualitativen und quantitativen Repräsentationen der Problemlösefähigkeit (Jonassen, 2003). Studierende sehen, dass Lernen ein kontinuierlicher Prozess von Interaktion zwischen Verstehen und Repräsentation ist. Sie professionalisieren ihr Wissen über das Subjekt und ihre RC. In diesem Prozess sind multimodale Repräsentationen eine förderliche Basis für die Erkenntnisgewinnung (Waldrup et al., 2010).

Die Fähigkeit MER zu erstellen, zu interpretieren, zu übersetzen und zwischen verschiedenen Darstellungen, zu verknüpfen wird Representational Fluency genannt. Die Representational Fluency, die angibt, wie gut die RC umgesetzt werden kann, wurde in den letzten Jahren vor allem in der Chemiedidaktik eingesetzt, um die Wichtigkeit des fließenden Übersetzens von Symbolischem, Makroskopischem und von Repräsentationen auf dem Teilchenlevel hervorzuheben (Chang, 2018).

Die Auseinandersetzung mit Repräsentationen und das Lernen mittels Repräsentationen, kann Lernende bei der Erkenntnisgewinnung unterstützen, indem sie Repräsentationen als förderliche Basis für die Erkenntnisgewinnung nutzen. Die RC umfasst die Kompetenzbereiche der Erkenntnisgewinnung und ebenso der Kommunikation der deutschen Bildungsstandards (Kultusministerkonferenz, 2005a, 2005b, 2005c) (siehe Kapitel 2). Repräsentationen sind ein wichtiges Werkzeug der Erkenntnisge-

winnung, da durch sie experimentelle Handlungen festgehalten und daraus Ideen zur Planung und Umsetzung entwickelt werden können. Kommunikation ist für Lernende ein notwendiges Werkzeug, um Erklärungen für Phänomene zu entwickeln, diese in geeigneter Form darzustellen (verbal, symbolisch, mathematisch) und mitzuteilen (Kultusministerkonferenz, 2005b).

3.4 Skizzieren

Das Skizzieren wird sowohl für das Herstellen von Skizzen und Zeichnungen im Sinne von Kunst wie auch jegliche Art von bildnerischen Darstellungen verwendet. Unter Zeichnen und Skizzieren wird, im Gegensatz zum Malen, das Herstellen von bildlich Darstellungen, die nur aus Linien und Strichen bestehen, verstanden. Beim Malen entsteht ein Bild aus Flächen. Das sogenannte *wissenschaftliche Zeichnen* bezeichnet in der Biologie das möglichst originalgetreue Abzeichnen von Objekten mit Hilfe von Lineal und anderen Hilfsmitteln (Honomichl et al., 2013). Während man beim Zeichnen besonderen Wert auf Details, Texturen und Schattierungen legt, geht es bei einer Skizze nur um eine schnelle Momentaufnahme. Formen und Proportionen sollen in kurzer Zeit mit geringem zeichnerischem Aufwand richtig erfasst werden. Skizzen werden oftmals ohne Hilfsmittel angefertigt. Nachfolgend werden die Begriffe *wissenschaftliche Skizze* und *wissenschaftliches Skizzieren* definiert und in den Oberbegriff des Dokumentierens eingeordnet (siehe Kapitel 3.4.1). Des Weiteren werden die allgemeinen Funktionen von *wissenschaftlichen Skizzen* dargelegt (siehe Kapitel 3.4.2), und im Fazit wird auf spezifische Anwendungszwecke des Skizzierens im naturwissenschaftlichen Unterricht eingegangen, die dann in die Kompetenzbereiche Kommunikation und Erkenntnisgewinnung (Kultusministerkonferenz, 2005b) eingeordnet werden (siehe Kapitel 3.4.3).

3.4.1 Beschreibung Skizzieren, Einordnung wissenschaftliches Skizzieren

Dokumentiert werden kann in unterschiedlichen Formen: in verbal-sprachlicher Form (Protokollieren), in nicht verbal-sprachlicher Form oder in Kombinationen beider Formen. Bei der nicht verbal-sprachlichen Form steht vor allem die *wissenschaftliche Skizze* (zeichnerische Darstellung) im Fokus. Daneben kommen alternative Formen wie Fotografie und Video zum Einsatz (Groß & Reiners, 2012). Cooper et al. (2017) sehen aber auch chemische Zeichnungsarten, wie bspw. die Lewisformeln oder Formeleinheiten, als *wissenschaftliche Skizzen*, nicht jedoch Rechnungen, Tabellen oder

Diagramme. Skizzen können durch verbal-sprachliche Teile oder mit symbolischen Formen, etwa Tabellen, Graphen, Schemata, Diagramme und Rechnungen, ergänzt werden. Kombinationen von Skizzen, Texten oder anderen grafischen Elementen werden „Multiple External Representations“ (MER) (siehe auch Kapitel 3.3) genannt. Darunter versteht man Graphen, Tabellen, Skizzen mit Legenden oder mit erklärendem Text sowie grafischen Elementen wie Diagramme und bildliche Darstellungen (Ainsworth, 1999; Pande & Chandrasekharan, 2017), die sowohl Bild (Skizze), Symbole als auch Text enthalten können (Chang, 2018). Concept Cartoons zählen ebenfalls dazu (Arnold et al., 2016). Wird eine Skizze durch andere Elemente ergänzt, entsteht eine MER. Abb. 1 zeigt einen Überblick über die Formen von Dokumentationen.

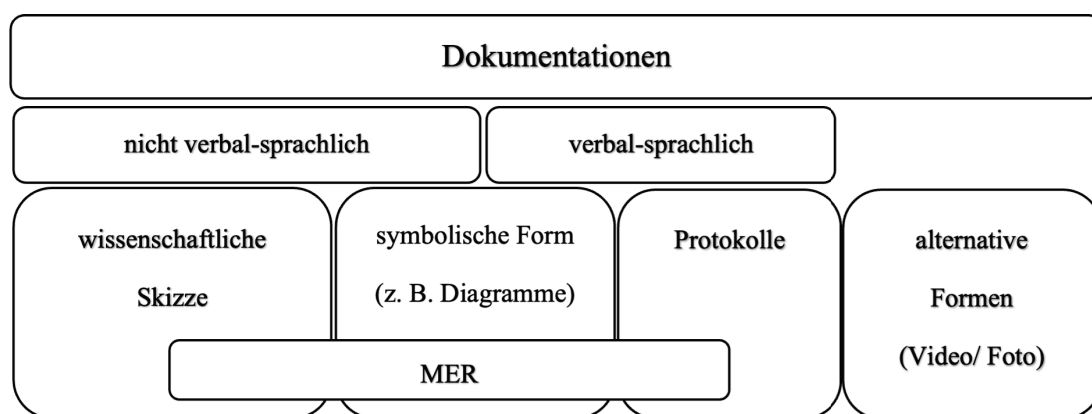


Abb. 1 Formen von Dokumentationen

Skizzieren wird in der Literatur unterschiedlich aufgefasst: Nach Ainsworth et al. (2011, S. 1096) entspricht das Skizzieren der Anfertigen von Darstellungsformen der Naturwissenschaften („Drawing to learn to represent in science“) und gilt als Repräsentationskonstruktion (Cooper et al., 2017). In Bezug auf das Lernen und Unterrichten in Chemie gibt es vier Typen von Repräsentationen, nämlich die symbolische, die systematische, die partikulare und die makroskopische Repräsentation (Gilbert, 2008; Lehrer & Schauble, 2006).

Skizzieren („Sketching“) wurde von Quillin und Thomas (2015) als „a learner-generated external visual representation depicting any type of content, whether structure, relationship, or process, created in static two dimensions in any medium“ definiert. Van Meter und Garner (2005) bezeichnen Schülerskizzen als bildliche Repräsentationen, welche absichtlich konstruiert sind, um ein Lernziel zu erarbeiten. Für deren Erstellung, die akkurate Abbildung des Objekts und für das Erscheinungsbild sind vor

allem die Lernenden zuständig. Causey (2016) sieht im Skizzieren einen Prozess, dessen Produkt das Sehen ist, im Sinne von „you are drawing to see, not seeing to draw“. Ähnlich betrachtet es auch Ingold (2011). Er beschreibt das Skizzieren als Spur der beobachtenden Handbewegung, die dem, was passiert, folgt. Mäkelä, Nimkulrat, and Heikkinen (2014) benennen das Skizzieren als ein aktives wie auch subjektives Engagement, welches Zugang zum persönlichen Wissen ermöglicht. Neben Skizze bzw. Zeichnung und dem Skizzieren bzw. Zeichnen werden in der Literatur weitere Begriffe verwendet, die den in dieser Arbeit verwendeten Begriff des *wissenschaftlichen Skizzierens* einschließen. Sogenannte „Inscriptions“ werden in Bezug auf Notationen, Skizzen, Graphen, Diagramme, Tabellen und andere visuelle Darstellungen verwendet (Greeno & Hall, 1997; Medina & Suthers, 2013).

Dem Wissen, wie Erkenntnisse gewonnen werden, und dem Können, durch Handlung Erkenntnisse zu gewinnen, wird im naturwissenschaftlichen Unterricht ein zentraler Stellenwert zugeschrieben. Erkenntnisse können durch experimentelles Handeln gewonnen werden (siehe Kapitel 3.2.1). Lernende halten ihre Handlungen und Ergebnisse beim Experimentieren verbal-sprachlich protokollierend und zeichnerisch fest: Sie dokumentieren diese. Bayrak und Ralle (2016) wie auch Engl und Risch (2015) sehen im Dokumentieren eine wissenschaftliche Tätigkeit, die als förderliche Basis für Erkenntnisgewinnung dient.

Durch Dokumentationen werden Kommunikation und Austausch über die Beobachtungen, Durchführung, Resultate und die daraus resultierende Erkenntnisgewinnung möglich. Das Dokumentieren in Form von Protokollieren ist breiter Forschungsgegenstand (Bayrak & Ralle, 2016; Engl & Risch, 2015), wohingegen das *Skizzieren als Teilbereich* des Dokumentierens kaum erforscht ist (Nitz et al., 2014).

Frischknecht-Tobler und Labudde (2010) betonen ebenfalls, dass sich das Skizzieren auf der Ebene des naturwissenschaftlichen Unterrichts nicht nur auf das (Ab)zeichnen von Objekten im Rahmen von Betrachtungen oder Beobachtungen beschränkt, sondern es in sämtlichen Phasen des Forschungszyklus unterstützt. Das Skizzieren bezeichnet dabei die Teilfähigkeit beim Dokumentieren, Skizzen während des Experimentierprozesses anzufertigen. Skizzieren ist also eine wissenschaftliche Tätigkeit, wie auch Protokollieren (Bayrak & Ralle, 2016; Engl & Risch, 2015), das eine förderliche Basis für den Erkenntnisgewinnungsprozesses ist, und findet sich grundsätzlich in allen Teilprozessen des Experimentierens wieder (Gut & Mayer, 2018). Durch

das Skizzieren werden diese Teilprozesse festgehalten. Zu diesen gehören die Entwicklung der Fragestellung, das Generieren der Hypothese, die Planung und Durchführung des Experiments, Messungen und Beobachtungen, das Ziehen von Schlussfolgerungen und die Reflexion (Gut & Mayer, 2018). Durch die Auseinandersetzung mit dem Forschungsgegenstand beim Festhalten und durch die Betrachtung der Skizzen können Erkenntnisse gewonnen werden.

Cooper et al. (2017) unterteilen Skizzen in drei Bereiche: Sogenannte *voraussagende (predictive)*, *beobachtende (observational)* und *reflektierende (reflective)* Skizzen. Dabei werden alle Schritte der experimentellen Handlung durchlaufen und in Skizzen festgehalten. Die *voraussagende Skizze* fokussiert den Bereich der Fragestellung, der Hypothesenbildung wie auch der Planung der Experimente. Mit der *beobachtenden Skizze* halten die Lernenden alle experimentellen Schritte und Beobachtungen (beispielsweise von Phänomenen oder den Vergleich von mehreren Phänomenen), den Aufbau und die Durchführung eines Experimentes fest. Die *reflektierende Skizze* wird nach der Experimentierphase hergestellt und zeigt die Auswertung und Beantwortung der Fragestellung.

Im weiteren Verlauf dieser Arbeit wird der Begriff *wissenschaftliche Skizze* in der Definition von Quillin und Thomas (2015) und van Meter und Garner (2005) zur Dokumentation von experimentellen Handlungen wie Beobachtungen, Planung und Durchführung von Arbeitsschritten verwendet. Dabei wird der Ansatz von Causey (2016) und Mäkelä, Nimkulrat und Heikkinen (2014) eingeschlossen, die Skizzieren als einen Prozess des Sehens und einen Zugang zum persönlichen Wissen auffassen. Skizzen sind von Lernenden eigenhändig und ohne Hilfsmittel, wie beispielsweise Lineal, erstellte Darstellungen.

Die breite Nutzung von Skizzen sowie verschiedene Funktionen, die Skizzen und Zeichnungen allgemein im schulischen und universitären Kontext erfüllen, werden im Kapitel 3.4.2 besprochen.

3.4.2 Funktionen des Skizzierens

Skizzen nehmen in schulischen und universitären Kontexten des naturwissenschaftlichen Unterrichts unterschiedliche Funktionen wahr. Sie dienen zum einen als Methode, um bspw. den Lernzuwachs zu steigern oder mentale Modelle zu aktivieren und zu entwickeln. Zum anderen dienen Skizzen als Erhebungsinstrument zur Ein-

schätzung kognitiver Kompetenzen oder zur Erhebung von Fachwissen. Außerdem stellt das Skizzieren eine Schülerkompetenz dar (siehe auch Kapitel 4). Im Folgenden werden die unterschiedlichen Funktionen aufgezeigt, und es wird auch eine klare Eingrenzung und Abgrenzung des Begriffs *wissenschaftliches Skizzieren* vorgenommen.

3.4.2.1 *Skizzieren als Methode*

Durch Skizzieraufträge mit Instruktionen können didaktische Ziele verfolgt werden. Diese führen Quillin und Thomas (2015) unterscheiden dabei zwischen formativen und summativen Sequenzen. Formative Übungen dienen dazu, dass Studierende mit Hilfe von Rückmeldungen ihr Wissen konstruieren und Fähigkeiten erlernen sowie dadurch das Verstehen und die Performance verbessern. Bei bildlichen Darstellungen steht die Förderung von aktivem Lernen, der Beobachtungsfähigkeiten, des Memorierens, Verstehens räumlicher Beziehungen und die Freude am Lernen im Zentrum. Bei abstrakten Skizzierungsübungen geht es zudem nicht nur um die Förderung der Konstruktion von mentalen Modellen und von Verknüpfungen von Ideen und Konzepten, sondern auch um die Aneignung von Kontextwissen, um Problemlösefähigkeit, um Prozessfähigkeiten in den Naturwissenschaften und nicht zuletzt um das Aufdecken von Fehlkzepten. In formativen Sequenzen steht das Aufdecken von Wissen, von mentalen Modellen, von Prozessfähigkeiten, von Problemlösefähigkeiten, von Inhaltswissen und die Aufzeichnung von Beobachtungen im Zentrum.

Als lernfördernde Methode findet sich das Skizzieren in verschiedenen Kontexten zur Kompetenzförderung und Wissensvermittlung: Mittels Skizzieren und anderen Visualisierungstätigkeiten werden Problemlösefertigkeiten (Cox, 1999; Ridley & Rogers, 2010; Roam, 2013) und Prozessfähigkeiten des Experimentierens (Quillin & Thomas, 2015) erlernt und gefördert. Skizzieren fördert das aktive Lernen wie auch das Verstehen von räumlichen Zusammenhängen und stellt einen Weg des visuellen Lernens dar (Quillin & Thomas, 2015). Als Lernmethode unterstützt das Skizzieren die Transformation von oberflächlichem Wissen in vertieftes Fachwissen (Kozma et al., 2000), da das Skizzieren Informationen vereinfacht (Quillin & Thomas, 2015) und verbessert (Lehrer & Schauble, 2006) und dadurch ebenso das Verständnis für Skizzen vereinfacht bzw. verbessert wird (Kiyokawa et al., 2012). Ferner fördert das Skizzieren die Erinnerung (Quillin & Thomas, 2015; Ridley & Rogers, 2010) und hilft damit bei der Aneignung von Fachwissen, zum Beispiel von biologischen Prozessen. Die Anfertigung von Skizzen fördert das vernetzte Denken und hilft beim Verstehen von

Fachinhalten (Britton & Wandersee, 1997), da zeichnerisch Informationen neu zugeordnet und dadurch Probleme gelöst werden können (Cox, 1999). Skizzieren gilt als eine Lernstrategie (Ainsworth et al., 2011; Quillin & Thomas, 2015), die das aktive Lernen fördert (Quillin & Thomas, 2015), weil Denkprozesse durch das Skizzieren angeregt werden. Während des Skizzierens entwickeln sich die Gedanken weiter (Ridley & Rogers, 2010; Wendell & Rogers, 2013), ähnlich wie dies auch beim Schreiben vom Versuchsprotokollen der Fall ist.

Cooper, Stieff und DeSutter (2017) wie auch Kozma und Russell (2005) schlagen deshalb vor, man solle modellbasiertes Denken in der Chemie mit Skizzieren fördern und das Skizzieren nicht als isolierte Aktivität, sondern als Teil unterschiedlicher Aktivitäten im Forschungszyklus verstehen und damit aktiv den Weg der Erkenntnisgewinnung unterstützen. Skizzieren ist ein wirkungsvolles Werkzeug für das Denken (Quillin & Thomas, 2015), weil es hilft, über das eigene Lernen nachzudenken (Ridley & Rogers, 2010). Quillin und Thomas (2015) empfehlen in ihrem Draw-to-Learn Framework ebenfalls, in den Naturwissenschaften mit Skizzieren modellbasiertes Denken oder das modellbasierte Argumentieren (Reasoning) zu betreiben (Cooper et al., 2017; Cox, 1999; Lehrer & Schauble, 2006; Quillin & Thomas, 2015). Weil Skizzieren oftmals den Kern chemischen Denkens darstellt (Kozma et al., 2000), würden ohne das Skizzieren Experten nicht in der Lage sein zu argumentieren (Kozma et al., 2000; Pande & Chandrasekharan, 2017). Skizzieren ist damit gemäß Ainsworth (2011, S. 1096) eine Form des Argumentierens („Drawing to Reason in Science“) und, wie Cox (1999) hervorhebt, zentral für das Argumentieren, denn nur eine Minderheit kann Probleme allein im eigenen Kopf (internes Modell) lösen. Während der Konstruktion von Repräsentationen und dem Skizzieren werden eigene Ideen überprüft, Informationen umgeordnet und von einer Modalität in die andere umgeformt sowie der Prozess der Problemlösung festgehalten (Cox, 1999). Die aktive Konstruktion von Skizzen und Repräsentationen ist wichtig für die Förderung des Verstehens der Naturwissenschaften (Ainsworth et al., 2011; diSessa, 2004; Greeno & Hall, 1997; Nitz et al., 2014).

Giere and Moffatt (2003) empfehlen Lernenden, Repräsentation als Werkzeug des Denkens für das Voraussagen, das Verständnis und das Aufstellen von Behauptungen zu nutzen, anstatt nur die korrekte Skizze für eine Wissensdarstellung zu memorieren. Die Fähigkeit, wissenschaftliche Phänomene zu verstehen (Cooper et al., 2017; Koz-

ma et al., 2000; Quillin & Thomas, 2015; Ridley & Rogers, 2010; Schwarz et al., 2009) und zu erklären (Kozma et al., 2000) wird durch das Skizzieren verbessert. Durch das Skizzieren kann „Unsichtbares“ sichtbar gemacht (Cooper et al., 2017; Schwarz et al., 2009) und wie ein System abstrahiert (Kozma et al., 2000) werden. Lernende können durch das Skizzieren Hypothesen bilden und Voraussagen zu einem Experiment machen (Cooper et al., 2017; Roth & McGinn, 1998). Prozessfähigkeiten (Teilprozesse des Experimentierens) können durch das Skizzieren erworben und gefördert werden (Quillin & Thomas, 2015). Skizzieren kann also als förderliche Basis für den Weg der Erkenntnisgewinnung angesehen werden.

Skizzen sind ebenso ein Weg, um etwas über mentale Modelle (Cox, 1999; Kelly et al., 2010; Martschinke, 1996; Roth & McGinn, 1998; Schwarz et al., 2009; Thomas et al., 2001) und Fehlkonzepte (Chi et al., 1981) zu erfahren. Mit den Untersuchungen von „Drawing a Man“ und „Draw-A-Science-Teacher“ (Finson, 2002; Thomas et al., 2001) wurden Präkonzepte und Modelle Lernender anhand ihrer Skizzen untersucht. Repräsentationen und Skizzen vermitteln zwischen Sichtbarem und Unsichtbarem (Kozma et al., 2000) und unterstützen dadurch die Entwicklung von Visualisierungsfähigkeiten (Hope, 2008). Als Visualisierungswerkzeug (Stieff, 2011) helfen Skizzen bei der Problemlösung, bspw. in der organischen Chemie (Stieff, 2011) oder bei der Ideenfindung durch Skizzieren im Allgemeinen (Kelley & Sung, 2017; Roam, 2013). Das Skizzieren stellt einen Weg des Modellierens dar, mit dem mentale Modelle und mentale Zustände entwickelt und dargestellt werden. In der Entwicklung von Repräsentationen können Phänomene dargestellt und damit verstanden werden (Lehrer & Schauble, 2006). Stein und Power (1996) beschreiben Einsatzmöglichkeiten des Skizzierens in der Schule: Der Lehrer lässt Lernende skizzieren, was sie sähen, wenn sie Wasser durch ein superstarkes Mikroskop betrachten würden. Im Unterrichtsverlauf werden weitere Skizzen angefertigt, wobei sich die zunehmend verbesserte Qualität der Skizzen auf den Lernerfolg auswirkt. Dabei dienen Skizzen als Hilfsmittel zum Lernen und ebenso als Möglichkeit, Wissen und wesentliche Daten, sogenannte „Key Data“ darzustellen (Ridley & Rogers, 2010). Nebenbei bieten sie die Möglichkeit, vorhandene Fehlkonzepte aufdecken.

Durch Skizzieren wird eine der Darstellungsformen der Naturwissenschaften erlernt („Drawing to Learn to Represent in Science“) (Ainsworth et al., 2011, S. 1096), denn Wissenschaftler benutzen Repräsentation, um wissenschaftliche Phänomene zu er-

klären (Kozma et al., 2000). Das Skizzieren gehört deshalb zu den Kernaktivitäten auf allen Niveaus der Chemieausbildung (Cooper et al., 2017; Stieff et al., 2010).

Das Denken wird eingeschränkt durch die Konstruktion der Sprache (Kozma et al., 2000), denn wir können sprachlich nur so denken, wie die Sprache konstruiert ist (Crosland, 1962). Veränderungen, wie bspw. Temperaturänderungen, können durch Sprache allein schlechter dargestellt werden. Durch Repräsentationen, wie Diagramme, sind Veränderungen deutlicher darstellbar (Bowen & Roth, 2002; Kozma et al., 2000). In der visuellen Darstellung von Experimentiervorgängen liegt großes Potential, einerseits für jüngere Kinder, deren Wortschatz noch nicht ausgeprägt ist oder die noch nicht (gut) schreiben können, andererseits jedoch auch für nicht sprachaffine oder aus einer anderen Sprachregion oder -kultur stammende Kinder und Jugendliche. Skizzieren kann auch über Kulturen hinweg (Cross-Culture) zur Kommunikation eingesetzt werden, wo durch Sprachschwierigkeiten Missverständnisse entstehen könnten (Ridley & Rogers, 2010). Bilder und Skizzen helfen, über wissenschaftliche Phänomene und Erforschungen zu kommunizieren (Cooper et al., 2017; Kozma et al., 2000; Lehrer & Schauble, 2006; Quillin & Thomas, 2015; Ridley & Rogers, 2010). Das Kommunizieren ist auf diese Weise mit dem Dokumentieren verknüpft, d.h. die Kommunikation findet über Dokumentationen statt (Ainsworth et al., 2011; Kozma et al., 2000). Außerdem zeigen etliche Forschungsergebnisse (Van Meter & Garner, 2005), dass Skizzen das Textverständnis verbessern und den Schreibprozess fördern. Das Skizzieren kann darüber hinaus das Lernen aus naturwissenschaftlichen Texten fördern (Ainsworth et al., 2011; Van Meter, 2001). Mehr noch: es ist allgemein ein wirksames und disziplinenunabhängiges Instrument der Kommunikation (Quillin & Thomas, 2015) und damit die Kommunikation- und Argumentationsfähigkeit (Cooper et al., 2017).

Das Skizzieren verbessert die Kompetenz, räumliche Zusammenhänge zu erfassen und zu verstehen (Quillin & Thomas, 2015). Dreidimensionales Denken kann durch Abzeichnen von Fotos und Skizzen sowie durch computerbasiertes Skizzieren (CogSketch) gefördert werden (Jee et al., 2014). Das Skizzieren und andere visuelle Darstellungsprozesse verbessern außerdem den Beobachtungsprozess, weil für die Darstellung die Vorgänge präzise beobachtet werden müssen. Das Beobachten (di-Sessa & Sherin, 2000) und die Festlegung des Fokus beim Beobachten (Cooper et al., 2017) werden gefördert. Die Betrachtung und die Dokumentation von Demonstra-

tionen chemischer Prozesse auf der Makro- und Submikroebene dient der Weiterentwicklung der Vorstellungskraft (Cooper et al., 2017). Dempsey und Betz (2001) lassen Lernende Details von Gegenständen aus der Erinnerung zeichnen und Skizzen auf Grund von Beschreibungen eines Mitlernenden anfertigen. Bedeutsam ist dabei auch die verbale Auseinandersetzung. Dadurch lernen Studierende, Bedingungen und Eigenschaften zu beschreiben.

Skizzieren fördert (Cooper et al., 2017) und schärft die Beobachtungsfähigkeit (diSessa & Sherin, 2000), also das systematische Beobachten, was in den Teilprozessen des praktischen Arbeitens während des Experimentierprozess im Sinne der Erkenntnisgewinnung zielführend eingesetzt werden kann. Die Entwicklung von Repräsentationskompetenzen (siehe Kapitel 3.3), wie bspw. das Interpretieren von Darstellungen oder die Fähigkeit, eigenständig adäquate Skizzen zu erzeugen (Engl et al., 2015; Schumacher & Roth, 2014), erhöht bei Jugendlichen das Verständnis für die Wissenschaften (diSessa, 2004) und fördert dabei ebenfalls die Fähigkeit der Erkenntnisgewinnung.

Das Anfertigen von Skizzen hat einen positiven Einfluss auf die Motivation im Lernprozess und steigert es das Interesse an der Vertiefung eines Themas (Ainsworth et al., 2011; Azevedo, 2000; Quillin & Thomas, 2015; Van Meter & Garner, 2005). Zudem kann durch Skizzieren besser memoriert (Ainsworth et al., 2011; Azevedo, 2000; Cox, 1999; Quillin & Thomas, 2015; Ridley & Rogers, 2010; Van Meter & Garner, 2005) und gemäß Quillin und Thomas (2015) kann der Overload an Informationen aufgenommen und verarbeitet werden, was wiederum die Motivation für den Lernprozess steigert.

Beim Skizzieren, das als Methode dient, handelt es sich nicht zwangsläufig um *wissenschaftliche Skizzen*, wenn die Dokumentation einer experimentellen Handlung und die damit verbundene Erkenntnisgewinnung nicht im Fokus steht. Es kann sich um eine *wissenschaftliche Skizze* handeln, falls die Skizze eine experimentelle Handlung festhält, aus der Erkenntnis gewonnen werden kann.

3.4.2.2 *Skizzen als Messart*

Das Mittel, das zur Messung einer Leistung oder Performance eingesetzt wird, wird als Messart oder Messmethode bezeichnet. Im Folgenden wird dargelegt, wie und wozu Skizzen als Messmethode in der Praxis eingesetzt werden. Skizzen als Messart

kommen zum Beispiel in Performance Assessments wie TIMSS (Martin et al., 2012) und ExKoNawi (Bonetti et al., 2017, 2018; Gut et al., 2014; Gut & Mayer, 2018) vor, um Kompetenzen wie das Experimentieren oder naturwissenschaftliche-technisches Handeln (Practical Work) als Ganzes zu erheben.

Skizzen von Lernenden zeigen einerseits deren Beobachtungen und andererseits Bestandteile ihres Wissen, weshalb Skizzen auch zur formativen und summativen Beurteilung genutzt werden können (Ainsworth et al., 2011; Quillin & Thomas, 2015). Chemische Bildersuche deckt bspw. vorhandenes ebenso wie fehlendes Fachwissen auf (Nickel & Engel, 2018). Bei einer Chemie-Foto-Story (Precht, 2007, 2008, 2013; Tomcin & Reiners, 2009), bei der Experimentiersituationen als eine Art Comics dargestellt werden, dient das Skizzieren mitunter auch als Analysewerkzeug der Konzepte der Lernenden (Finson, 2002; Thomas et al., 2001). Abstrakte Skizzen hingegen vermögen zudem mentale Modelle (Cox, 1999; Martschinke, 1996; Roth & McGinn, 1998; Schwarz et al., 2009; Thomas et al., 2001) und Fehlkonzepte dazustellen wie auch das Zusammenhangswissen von Daten und die Problemlösefähigkeit (Quillin & Thomas, 2015) zu unterstützen. Dies kann zur Diagnose benutzt werden.

Gemäß Quillin und Thomas (2015) können durch Skizzieren Prozessfähigkeiten (Process Skills) aufgedeckt werden, beispielsweise Hypothesen bilden, Experimente entwerfen, Visualisieren und Interpretieren sowie Resultate kommunizieren. Des Weiteren kann mit Skizzen das Lernen der Schülerinnen und Schüler gemessen werden, sei es die Erfassung von tieferem Verständnis (Ainsworth, 1999), die Verknüpfung zwischen Repräsentationen oder ob sie Informationen von einer Repräsentation in eine andere übertragen können. Kodierungen von Schlüsselmerkmalen in Skizzen (Jee et al., 2014) oder Klassifizierungen (Solano-Flores et al., 1997) von dargestelltem Fachwissen können für die Beurteilung eingesetzt werden.

In Experimentierphasen werden durch die Lernenden neben Protokollen (sprachlich-verbale Dokumentation) auch Skizzen oder Repräsentationen, wie etwa Diagramme (nicht sprachlich-verbale Dokumentation), angefertigt. Oftmals werden die Lernenden durch einen konkreten Auftrag („zeichne“) oder eine leere Seite (anstelle von Linien) zum Skizzieren angeregt.

Experimentiersammlungen wie die von Lunetta und Novick (1982) enthalten neben Experimenten oftmals auch Anregungen zur Aktivierung von Lernenden. Diese werden animiert, selbst Erkenntnisse zu gewinnen, indem sie bspw. zum „Diskutieren

oder Zeichen eines Graphen“ aufgefordert werden. Skizzen dienen in diesem Zusammenhang der Förderung des Erkenntnisgewinns bzw. des Lernzuwachses von Fachinhalten. Es handelt sich bei den Skizzen um MERs, jedoch selten um *wissenschaftliche Skizzen*. Zudem werden Art und Qualität der Skizzen dabei nicht berücksichtigt.

In Tests zur Erhebung von unterschiedlichen Kompetenzen spielen Skizzen eine Rolle. Sie stellen in Hands-On Assessments und -Tests neben Text und Multiple-Choice ein mögliches Antwortformat (Brown & Shavelson, 1996) dar. Freie Flächen mit der Anweisung wie „Halte die Dinge, die du tust und beobachtest, fest.“ oder „Schreibe in dein Laborjournal.“ in Hands-On Assessments bieten die Möglichkeit, zeichnerisch Gedanken zu entwickeln und festzuhalten. Skizzen dienen vor allem als Hilfsmittel, Experimentieranleitungen darzustellen. Explizite Anweisungen zum Skizzieren gibt es keine. Die Skizzen sind ein Werkzeug, um Prozesse (Handlungskompetenzen) oder Fachwissen (bspw. Schaltschemen von Stromkreisen) abzubilden. Neben dem überwiegenden Antwortformat Text dienen auch Repräsentationen in Form von Diagrammen oder Skizzen von Versuchsaapparaturen als mögliche Antworten in Hands-On Tasks zur Erfassung von Fachwissen (Bryce, 1984; Stecher & Klein, 1996). Jedoch boten Hands-On-Tests bislang keinen Zugang zur Kompetenz des *wissenschaftlichen Skizzierens* (Brown & Shavelson, 1996).

In der TIMSS (Harmon, 1997) werden Repräsentationen wie zum Beispiel Diagramme teilweise als Antwort erwartet. Skizzen werden aber explizit nur als Skizze einer Körperoberfläche eines Quaders erwartet. Skizzen oder Zeichnungen gelten laut Kodiermanual als ein mögliches Antwortformat, für die Bewertung jedoch ist nur die fachliche Korrektheit des Inhalts der Skizze relevant. Auch in den Studien des British Columbia Assessment of Science (Widden et al., 1992) und des Assessment of Performance Units Science at Age 13 (Schofield, 1989) wird bei Skizzen nur das auf Skizzen dargestellte Fachwissen in der Bewertung berücksichtigt. Für die Festlegung der Bildungsstandards der Schweiz (EDK Erziehungsdirektoren-Konferenz, 2011) werden Skizzen ebenfalls explizit miteinbezogen. Neben Multiple-Choice und kurzen Texten gelten auch Skizzen als mögliches Antwortformat. Lernende werden durch die Aufforderung „Zeichne oder beschreibe“ oder ganz konkret zum Beispiel durch „Zeichne einen Spiegel“ zum Erstellen von Skizzen angeregt (Wissenschaftliches Konsortium HarmoS, 2008). Skizzen werden nur auf die inhaltlich-fachliche Korrekt-

heit geprüft. Die Art der Skizzen wie auch die Rechtschreibung von Texten werden in der Bewertung nicht berücksichtigt.

Beim Projekt ExKoNawi (experimentelle Kompetenzen in den Naturwissenschaften) (Bonetti et al., 2017, 2018; Gut et al., 2014) handelt es sich um ein Hands-On Assessment, das experimentelle Kompetenzen nach dem Problemtypen-Ansatz modelliert (Gut et al., 2014). Die Antwortformate reichen von verbal-sprachlichen (z.B. Multiple-Choice) über kombinierte (Text/Skizze) bis hin zu reinen Skizzen. Bei den Skizzen wird die fachliche Korrektheit des Inhalts oder die korrekte fachmethodische Anwendung bewertet. Weder die Art der Skizze noch die *Wissenschaftlichkeit der Skizze* haben Einfluss auf die Bewertung.

Zusammenfassend lässt sich festhalten: Skizzen sind Bestandteil der Kompetenzmessung, jedoch ohne, dass es um die Kompetenz des *wissenschaftlichen Skizzierens* geht. Skizzen dienen zwar in verschiedenen Tests und Assessments als ein mögliches Antwortformat, jedoch wird weder die Art noch die Qualität der Skizze bewertet oder beurteilt. Die Skizzen dienen zur Messung von Fachinhalten, Fachwissen, fachmethodischen Anwendungen und zur Messung von Schülerkompetenzen, die durch die Anwendungen von Skizzen gefördert werden. Bei Skizzen, die als Messmethode genutzt werden, kann es sich sowohl um *wissenschaftliche Skizzen* handeln als auch um Zeichnungen, die nicht als *wissenschaftliche Skizze* bezeichnet werden können. Wird eine experimentelle Handlung mit dem Ziel der Erkenntnisgewinnung dokumentiert, handelt es sich um eine *wissenschaftliche Skizze*, welche bis anhin für die Messung von anderen Kompetenzen eingesetzt wird.

3.4.2.3 *Skizzieren als Schülerinnen- und Schülerkompetenz*

Die Fähigkeit verbal-sprachlich zu dokumentieren, die sogenannte Protokollierfähigkeit, ist verbunden mit der Kompetenz im Umgang mit einer vorgegebenen Darstellung (Engl et al., 2015) und gilt daher als eine Teilkompetenz der Representational Competence (RC) (siehe Kapitel 3.3). Engl und Risch (2015) erwägen, dass die Fähigkeit des Protokollierens eine wichtige Kompetenz im Prozess der Erkenntnisgewinnung darstellen könnte, da der Teilprozess gemäß Teilprozessansatz „Daten notieren und auswerten“ (Gut & Mayer, 2018) (siehe Kapitel 3.1.2) durch die Erstellung eines Protokolls gesichert wird und dadurch die förderliche Basis für die Erkenntnisgewinnung bieten. *Wissenschaftliche Skizzieren* deckt dieselbe Repräsentationsfunk-

tion ab und sind somit ebenfalls eine wichtige förderliche Basis für Prozess der Erkenntnisgewinnung. In der fachdidaktischen Forschung gibt es bisher keine Studien, die diese explizit untersuchen.

Kozma et al. (2000) zeigen, dass Naturwissenschaftlerinnen und Naturwissenschaftler über signifikante Repräsentationsfähigkeiten und -kompetenzen verfügen. Sie nutzen ihr Wissen, ihre Repräsentationsfähigkeiten und die Merkmale von Repräsentationen zum Aufstellen von theoriebasierten Gruppierungen von Problemlösungen und zur Argumentation ihrer Lösungen.

Hilfestellungen während des Skizzierens führen zu besseren Leistungen in Post-Test-Assessments (Alesandrini, 1981; Van Meter & Garner, 2005), da sie vermehrte Selbstkontrollen der Lernenden fördern. Lernende verbessern dadurch häufiger Verständnisfehler als sie das ohne Hilfestellungen tun würden (Van Meter, 2001).

Als Fördermaßnahme wurde ein Hilfsmittel zur Einführung des Dokumentierens als Methode zur Förderung des wissenschaftlichen Skizzierens in der Sekundarstufe I entwickelt (Cavelti et al., 2018). Der sogenannte Dokumentier-Führerschein umfasst neben der Einführung ins *wissenschaftliche Skizzieren* angelehnt an die Aufgabenkultur von Cooper (2017) auch die Schreibförderung durch Textbausteine und Ansätze zur Erkenntnisgewinnung mittels einem halboffenen Experiment.

Wenig ist bekannt über das Lernen, welches stattfindet, wenn Lernende Repräsentationen (Diagramme) skizzieren, um ihr Wissen darzulegen (Tippett, 2016). Plausibel erscheint jedoch, dass das erforderliche Inhaltswissen zusammen mit der Zeichnungsaktivität entwickelt wird. Unklar hingegen ist, wie die Studierenden das Inhaltswissen nutzen, um eigene Repräsentationen zu erstellen (Medina & Suthers, 2013). Weil sich Wissen und Fähigkeit während des Skizzierens entwickeln, impliziert dies, dass die Kompetenz, eigene Repräsentationen herzustellen, kontextabhängig sein könnte (Chang, 2018).

3.4.3 Fazit Skizzieren

Skizzen kommen in verschiedenen Kontexten und zu verschiedenen Anwendungszwecken zum Einsatz: als Methode, als Messart und als Kompetenz. Skizzieren kann das Argumentieren mit Modellen als Weg der Erkenntnisgewinnung (Kultusministerkonferenz, 2005b) unterstützen, indem Argumentationen zeichnerisch entwickelt und dargestellt werden (Cooper et al., 2017).

Wissenschaftliches Skizzieren ist sowohl aus wissenschaftlicher Sicht, bspw. als Repräsentationsfunktion (Ainsworth, 1999), als auch für die Schulpraxis als Handlungskompetenz relevant, da Experimentierprozesse zum Beispiel nicht alleine verbal dargestellt werden können (Kozma et al., 2000). Klassische Dokumentier-Anlässe in der naturwissenschaftlichen Praxis finden sich in den Erkenntnisgewinnungsmethoden des Beobachtens, des Vergleichens und des Experimentierens. Bei Anlässen des Modellierens wird oftmals mit Skizzen als Darstellungsmethode gearbeitet, die Arbeit mit Modellen wird jedoch selten dokumentiert, da dies eine extreme Abstraktionsleistung der Lernenden erfordern würde. *Wissenschaftliches Skizzieren* dokumentiert experimentelle Handlungen wie Beobachten Vergleichen oder Experimentieren im Sinne der Variablenkontrollstrategie (VKS), jedoch nicht das Modellieren oder Argumentieren. *Wissenschaftliches Skizzieren* stellt die Teilfähigkeit dar, beim Dokumentieren Skizzen des Experimentierprozesses anzufertigen, welche die Grundanforderungen Eindeutigkeit, Idealisierung, Vollständigkeit, Abstraktion und Korrektheit erfüllen. Eine Skizze, die diese Grundanforderungen erfüllt, bietet die förderliche Basis für eine erfolgreiche Erkenntnisgewinnung und unterstützt diese. *Wissenschaftliches Skizzieren* als selbstständige Kompetenz wurde in der Vergangenheit nicht vollumfänglich definiert. Es gibt wenig empirische Befunde bezüglich des Kompetenzniveaus der Lernenden, und es fehlen validierte large-scale Messinstrumente zur Erfassung der Kompetenz. Aussagen über das *wissenschaftliche Skizzieren* waren bislang nur indirekt möglich, indem das Skizzieren bei Kompetenztests zum Experimentieren als Teilanforderung mitgemessen wurde.

Wissenschaftliches Skizzieren wurde außerdem bislang nicht gegenüber anderen Kompetenzen, wie der experimentellen Kompetenz oder der Repräsentationskompetenz, abgegrenzt. Die Kompetenz des *wissenschaftlichen Skizzierens* zur Dokumentation von experimentellen Handlungen, wie Beobachtungen, Planung von Arbeitsschritten, Durchführung und Auswertungen, als förderliche Basis der Erkenntnisgewinnung wird in Kapitel 4 definiert und gegenüber anderen Kompetenzen abgegrenzt.

3.5 Kompetenzmodelle

Durch Kompetenzmodelle lassen sich die Kompetenzanforderungen allgemein, aber auch die der Bildungsziele strukturieren. Kompetenzmodelle sind Hypothesen zu Strukturen und Progressionen und dienen dazu, die abstrakt formulierten Bildungsziele in konkrete Aufgabenstellungen zur Leistungsmessung umzuformen. Mittels

empirischer Überprüfung, Messung, Systematisierung und Strukturierung, wie im Falle von Leistungstests, kann die Ausprägung spezifischer Kompetenzen ermittelt werden (Ferber, 2014; Klieme et al., 2003). Es wird zwischen Kompetenzstrukturmodell und Kompetenzentwicklungsmodell unterschieden (Schecker & Parchmann, 2006). Kompetenzstrukturmodelle beschreiben die Kompetenzbereiche auf einer Stufe und geben Aufschluss über die Kompetenzausprägung eines Lernenden zu einem bestimmten Zeitpunkt. Kompetenzentwicklungsmodelle dienen dazu, die Entwicklung von Kompetenzen eines Lernenden über einen zeitlichen Verlauf zu erfassen (Schecker & Parchmann, 2006). Soll eine Kompetenz nicht nur beschrieben, sondern auch bewertet werden, ist die Modellierung der Progression erforderlich. Progressionsmodelle sind Hypothesen von Konstrukten und dienen der Beschreibung von Niveaus (Mayer & Wellnitz, 2014). Progressionsmodelle lassen sich auf zwei Arten modellieren: Einerseits über die Merkmale der Anforderungssituation, andererseits über die Merkmale der Bewältigung einer solchen Anforderungssituation (Gut & Mayer, 2018). Die Testkonstruktion ist stark an die Art des Progressionsmodells gebunden, sofern es beim Test um die Überprüfung des Modells geht. Bei einer Anlehnung an die Anforderungssituationen ist eine große Zahl von Items unterschiedlicher Schwierigkeitsgrade notwendig, um Vergleiche zwischen theoretischen angenommenen und tatsächlich gemessenen Schwierigkeiten zu modellieren. Bei a-priori-Modellierungen der Anforderungssituationen liegt der Fokus bei der Beurteilung der Qualität der Bewältigung der Anforderungssituation (Gut & Mayer, 2018). Die Einteilung der Kompetenzausprägungen ausgehend vom Kompetenzstrukturmodell in Progressionsstufen kann nach unterschiedlichen Aspekten erfolgen, bspw. durch die Beschreibung ansteigender Fähigkeiten von einer unsystematischen zu einer systematischen Vorgehensweise (Hammann, 2004) oder durch die zunehmende Komplexität des wissenschaftsmethodischen Anspruchs (Wellnitz, 2012).

Bei der Testentwicklung im Bereich von Performance Assessments werden zwei Vorgehensweisen bei der Entwicklung und Überprüfung eines Progressionsmodells oder eines Kompetenzmodells unterschieden. Ein a-priori-Kompetenzmodell wird aus theoretischen Überlegungen und basierend auf bekannter Literatur entwickelt. Es beschreibt sowohl die Struktur wie auch die Niveaus der zu messenden Kompetenz. Mit diesem Modell werden literaturbasiert Hypothesen über die Struktur und Progression einer angenommenen Kompetenz postuliert, die empirisch zu überprüfen sind. Dazu wird auf der Basis dieses Kompetenzmodells ein Itemkonstruktionsmodell abgeleitet

und ein Test konstruiert, mit dem die im Kompetenzmodell enthaltenen Hypothesen überprüft werden sollen. Beispiele dazu sind das Modell des Projekts experimentelle Kompetenzen in den Naturwissenschaften ExKoNawi (Bonetti et al., 2017; Gut et al., 2014) und ESNaS (Kremer et al., 2012b). Bei ExKoNawi wurden die Kompetenzstufen für unterschiedliche experimentelle Problemtypen theoriebasiert a-priori entwickelt (Bonetti et al., 2017; Gut et al., 2014, 2017). Im Fokus standen fachmethodische Kompetenzen wie die Variablenkontrolle.

Alternativ wird in der Forschung jedoch häufig nur von einem Strukturmodell der Kompetenz ausgegangen, auf dessen Basis ein Itemkonstruktionsmodell und ein Test entwickelt werden. Ein Beispiel dafür ist PISA 2000 (Baumert et al., 2001). Aufgrund der Analyse von Itemschwierigkeiten im Test werden post-hoc Niveaustufungen der Kompetenz formuliert. Dies gelingt aus Erfahrung, siehe bspw. PISA 2000 (Baumert et al., 2001), nur schlecht.

3.5.1 Kompetenzmodelle in Bezug auf das wissenschaftliche Skizzieren

In Bezug auf das *wissenschaftliche Skizzieren* als Schülerinnen- und Schülerkompetenz (siehe auch Kapitel 3.4) gibt es bislang in der Literatur kein spezifisches Modell, das die Struktur oder die Stufung einer entsprechenden Kompetenz beschreibt. Ansätze einer Kompetenzmodellierung in Bezug auf Zeichnungen von Lernenden weist Hope's Framework (2008) auf, ist das Framework jedoch darauf ausgerichtet Design-Entwürfe zu bewerten, die Kompetenz, Skizzen zu erstellen steht nicht im Fokus.

3.5.1.1 *Hope's Framework*

Gill Hope (2008) entwickelte ein Bewertungssystem für Skizzen von Kindern in Bezug auf das entwickelte Verständnis und auf ihre kognitiven Fähigkeiten in der Entwicklung von Ideen und deren Umsetzung. Sie unterteilt den Zweck der Skizze in zwei Kategorien. Skizzieren kann zum einen ein als sogenanntes „Thinking Tool“ (Denkwerkzeug) aufgefasst werden; darunter fallen die von Hope kategorisierten Skizzentypen „Picture“, „Single-draw“ und „Multi-draw“. Sie bezeichnet diese als „Drawing as Containers for Ideas“. Durch das Skizzieren können die Lernenden Design-Probleme lösen. Mit „Journey of Design Thinking“ umschreibt Hope (2008) einen weiteren Zweck des Skizzierens auf und bezeichnet ihn als „Plateau of Realiza-

tion“, wobei Skizzieren hier eine kognitive Fähigkeit (Skill) darstellt. Dazu zählen Hopes Skizzentypen der Kategorie „Multi-design“, „Progressive“ und „Interactive“.

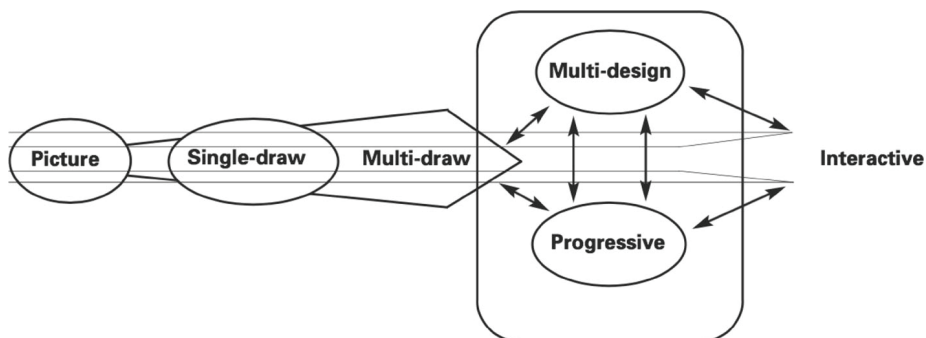


Abb. 2 entwickelndes Verständnis für das Skizzendesign in den Skizzentypen nach dem Framework von Hope (2008)

Hope (2008) untersuchte und klassifizierte Skizzen von Kindern im Vorschul- und Primarschulalter und stellte Kriterien auf, wie die Kinder das Skizzieren nutzen (siehe Abb. 2). Bei den Skizzentypen „Single-draw“ und „Multi-draw“ skizzieren die Kinder eine einzelne Idee ohne Weiterentwicklungs- oder Umsetzungsmöglichkeiten, zudem fehlt der Weg der Entwicklung dieser Idee. In die Kategorie „Multi-draw“ fallen die Skizzen, bei deren Anfertigung ein Kind zwar die perfekte Skizze anstrebt, jedoch eine einzelne Idee wiederholt und überarbeitet. Wenn ein Kind verstanden hat, dass skizzieren ein gutes Medium zur Weiterentwicklung von Ideen ist, bevor es das entwickelte Design umsetzt, dann gelingt es dem Kind, das Skizzieren für die Weiterentwicklung von Design-Ideen zu nutzen. Diese Formen nennt Hope (2008) „Multi-draw“ und „Progressive“. Als „Interactive“ bezeichnet sie den Punkt, an dem ein Kind ein Gespräch mit der Skizze führt. Es sieht die Skizze als Mittel auszudrücken, wie es vorgehen will und was zu tun ist. Das Kind ist in der Lage, durch Skizzieren und Schreiben Design-Ideen zu entwickeln.

Kelley und Sung (2017) untersuchen Design-Aufträge von Kindern in ihrer Studie nach dem Framework von Hope, indem sie eine ausführliche Kriterienliste für alle Kategorien entwickeln. Sie erkennen, dass nur wenige das „Plateau of Realisation“ erreichen. Sie stellen zudem fest, dass Kinder, die über eine gute Instruktion mit einigen grundlegenden Skizzierstrategien verfügen und Unterstützung durch die Lehrperson bekommen, ihre Skizzierfähigkeiten verbessern können.

Die Einteilung nach Hope (2008) stellt ein Kompetenzmodell dar, in welchem Progressionen modelliert werden. Die Einteilung der Niveaus entspricht den verschiede-

nen Stufen der Progression. Betrachtet wird jedoch nicht die Kompetenz des *wissenschaftlichen Skizzierens* (siehe Kapitel 4.1), sondern die Entwicklung von Designideen. Dennoch können die Ansätze des Framework von Hope (2008) für die Entwicklung des Kompetenzmodells des *wissenschaftlichen Skizzierens* (siehe Kapitel 6.1 und 6.2) angewandt werden.

3.5.2 Qualitätsmerkmale wissenschaftlichen Skizzierens

In der Literatur werden verschiedene Qualitätsmerkmale und Kriterien des Skizzierens wie auch für die Darstellung von Repräsentationen, die auch für die Kompetenz des *wissenschaftlich Skizzierens* Gültigkeit haben, diskutiert. Diese bilden die Grundanforderungen an *wissenschaftliche Skizzen* und somit für die Entwicklung eines Kompetenzstrukturmodells (siehe Kapitel 6.1). Die Qualitätsmerkmale und Kriterien werden im Folgenden aufgeführt.

DiSessa (2004) untersuchte naturwissenschaftliche Repräsentationen von Studierenden (z.B. Darstellung eines Bewegungsablaufs), während diese im Erstellen von Repräsentationen unterrichtet wurden, auf Entwicklungen und stellte folgendes fest:

- a) Figürliche Effekte: Studierende nutzen Muster, ohne dabei zu berücksichtigen, ob diese den systematischen Regeln entsprechen.
- b) Gewünschte Attribute: Studierende skizzieren sehr rasch die gewünschten Attribute, oftmals in einer Ad-hoc-Art, und umgehen dadurch die Systematik (Repräsentationsregeln).
- c) Realismus: Die Entwicklung der RC (siehe Kapitel 3.3) findet im Alltag statt (z.B. eine Geschichte erzählen); deshalb sind viele Repräsentationen realistisch gezeichnet. Bei wissenschaftlichen Repräsentationen werden Abstriche beim Realismus zugunsten der Systematik gemacht.
- d) Abstraktheit: Die Wissenschaft fokussiert sich mehr auf die Abstraktheit und indirekte Attribute und bewirkt so eine Verschiebung von Realismus zu Abstraktheit. Studierenden fällt der nichtwissenschaftliche Kontext jedoch leichter.
- e) Systematik: Studierende stellen die scheinbare Vollständigkeit (gefühlte und beobachtete) über die Systematik.
- f) Sorgfältigkeit: Studierende skizzieren oft intuitiv und sind nicht in der Lage, die Repräsentationsregeln zu erklären oder ihre Strategien zu begründen.

- g) Metrisches Verhältnis: Zahlen und numerische Verhältnisse sind in der Wissenschaft wichtiger als im Alltagsleben. Studierende lernen, diesen mehr Beachtung zu schenken und sie in die Skizzen einzubeziehen.

Zur Beurteilung von Skizzen der eindimensionalen Bewegung eines Fahrzeuges stellte diSessa (2004) folgende Kriterien auf:

- a) Vollständigkeit (sind alle relevanten Informationen dargestellt)
- b) Dichte (wird der Platz optimal genutzt)
- c) Präzision (sind die quantitativen Informationen präzise lesbar)
- d) Sparsamkeit (werden keine unnötigen Elemente genutzt)
- e) Systematik (werden einfache Regeln befolgt)
- f) Konventionen (werden keine Konventionen gebrochen)
- g) Erlernbarkeit (einfach zu erklären)

Zur Beurteilung von Skizzen dreidimensionaler Darstellungen stellte diSessa (2002) zudem folgende Kriterien auf:

- a) Erkenntnisbezogene Genauigkeit (Vollständigkeit und Genauigkeit)
- b) Systematik (einfache, einheitliche Regel)
- c) Eindeutigkeit (hat eine einzige Interpretation)
- d) Einfachheit (vermeidet unnötige Komplexität)
- e) Autonomie (aus eigener Kraft nachvollziehbar)
- f) Konventionen (werden keine Konventionen gebrochen)
- g) Anordnung (die Beziehungen zwischen den Teilen der Repräsentation sind klar dargestellt)
- h) Metrische Beziehungen

Azevedo (2000) untersuchte Repräsentationen von Landschaften mit dem Fokus der zweidimensionalen Darstellung des Dreidimensionalen. Kriterien sind der Einsatz von Farben, die Realitätstreue der Skizze, die konsequente Einhaltung der Beschriftung und die Frage, ob mehrere Betrachtungswinkel in die Skizze integriert wurden. Die Skizzen der Studierenden wurden kategorisiert nach der Qualität der Darstellung der Topografie in der Zweidimensionalität, wobei die Förderung der Studierenden in Bezug auf die Erstellung qualitativ besserer Skizzen im Fokus lag.

Jee et al. (2014) untersuchten, wie Laien (Studierende der Psychologie) und Studierende der Geologie geologische Strukturen zeichnen. Grundlage für die Skizzen waren eine Fotografie und ein Diagramm, die den Seduktions-Prozess der Plattentektonik darstellten. Die Kodierung der Skizzen erfolgte nach Hauptmerkmalen. Die Studierenden der Geologie zeichneten mehr Hauptmerkmale als die Laien. Auffällig war zudem, dass die Studierenden der Geologie in ihren Skizzen zusätzliche relevante Symbole verwendeten, die nicht in der Fotografie vorkamen, so z.B. Pfeile, um Bewegungsabläufe darzustellen (Jee et al., 2014). Auch in den Skizzen von Lernenden, die im Rahmen des Dokumentier-Führerscheins (Cavelti et al., 2018) angefertigt wurden, setzten die Lernenden Pfeile ein für die Darstellung von Bewegungen und Bewegungsabläufen oder, um eine Abfolge von Skizzen in eine Reihenfolge zu bringen. Unter einer sogenannten „Infografik“ versteht man, analog zu MERs, eine visuelle Repräsentation von Informationen und Daten, bei der durch die Kombination von Bildern, Text und Diagrammen komplexere Probleme beleuchtet und leicht verständlich geklärt werden. Durch die Auswertung von fast tausend Studierendenfeedbacks im Bereich der Infografiknutzung haben Gebre und Polman (2016) Aspekte für die Beurteilung von wissenschaftlichen Infografiken zusammengestellt. Die einzelnen Aspekte können auch für die Beurteilung einzelner Bestandteile der Infografik, also bspw. von Skizzen, genutzt werden. Die folgenden Aspekte sind relevant für die Auswertung einer Infografik:

- a) Klarheit (Aussagekraft)
- b) Vollständigkeit (abhängig von Daten, Inhalt, Umfang und Tiefe)
- c) Übereinstimmung zwischen den einzelnen Teilen
- d) Organisation und Design
- e) Sprache
- f) Lesbarkeit
- g) Relevanz
- h) Repräsentation
- i) Echtheit/Quelle der Ursprungsdaten

Eine Erklärung ist eine Verknüpfung von Beobachtetem mit eigener Interpretation, die vom fachwissenschaftlichen Wissen abhängig ist. Die Unterscheidung von Beobachtung und Erklärung ist Bestandteil des Kompetenzbereichs der Erkenntnisge-

winnung (Klos et al., 2008). Die Lernenden sollen Beobachtung und Erklärung klar trennen können und nur Beobachtbares zeichnerisch festhalten (Klos et al., 2008).

Zusammenfassend konnte gezeigt werden, dass die Kompetenz des *wissenschaftlichen Skizzierens* nur in geringem Masse im Fokus der naturwissenschaftlichen didaktischen Forschung liegt. Der Forschungsschwerpunkt liegt vielmehr in den Bereichen *Skizzieren* in Assessments oder *Skizzieren* als Methode, Fachinhalte oder andere Kompetenzen zu erwerben. Die Forschung im Bereich des *wissenschaftlichen Skizzierens* als förderliche Basis für die Erkenntnisgewinnung fehlt größtenteils. Ebenso gibt es weder eine präzise Definition des Konstrukts des *wissenschaftlichen Skizzierens* als förderliche Basis der Erkenntnisgewinnung experimentellen Handelns noch eine Abgrenzung und Einordnung zu anderen Konstrukten wie der experimentellen Kompetenz oder der Repräsentationskompetenz. Im Kapitel 4 werden Überlegungen zur Modellierung der Kompetenz des *wissenschaftlichen Skizzierens* wie auch Überlegungen zur Abgrenzungen gegenüber anderen Konstrukten besprochen.

4 Überlegungen zur Modellierung der Kompetenz des wissenschaftlichen Skizzierens

Skizzieren soll nach Frischknecht-Tobler und Labudde (2010) im naturwissenschaftlichen Unterricht in sämtlichen Phasen des Forschungszyklus stattfinden. Betrachtet man die matrixartige Verknüpfung der naturwissenschaftlichen Arbeitsweisen mit dem wissenschaftlichen Denken von Nehring et al. (2016) (siehe Tab. 2, Kapitel 3.2.1), kann Skizzieren in allen Feldern der Matrix die experimentelle Kompetenz unterstützen und dabei der Erkenntnisgewinnung dienen.

In der naturwissenschaftlichen Schulpraxis werden vor allem die naturwissenschaftlichen Arbeitsweisen (Nehring et al., 2016) Experimentieren und Beobachten, Vergleichen/Ordnen durch *wissenschaftliches Skizzieren* festgehalten. Durch Verknüpfung der naturwissenschaftlichen Arbeitsweisen mit dem wissenschaftlichen Denken (Nehring et al., 2016) halten wissenschaftliche Skizzen Teilschritte des wissenschaftlichen Denkens fest. Diese Matrixfelder können einzeln betrachtet werden: Bei der Verknüpfung der naturwissenschaftlichen Arbeitsweise „Experimentieren“ mit dem wissenschaftlichen Denken „Planung und Durchführung“ kann durch eine *wissenschaftliche Skizze* einerseits die Planung der VKS (siehe Kapitel 3.2.2.3) und andererseits die Durchführung derselben festgehalten werden. Die beiden Schritte Planung und Durchführung und die dazugehörigen *wissenschaftlichen Skizzen* gemäß dem Teilprozessansatz nach Gut und Mayer (2018) können auch einzeln betrachtet werden.

Wie in Kapitel 3.4.2 dargelegt, wird Skizzieren im Forschungsfeld im Allgemeinen breit gefasst, der hier definierte Begriff *wissenschaftliches Skizzieren* dagegen ist deutlicher enger und damit eingeschränkter. Für die Modellierung der Kompetenz des *wissenschaftlichen Skizzierens*, das die makroskopische Ebene abbildet und Vor- und Fachwissen ausschließt, sind nicht alle Felder der Matrix gleichermaßen geeignet. *Wissenschaftliche Skizzen* dokumentieren nur die Makroebene, also die Oberflächenstruktur von Phänomenen. Voraussagen, Hypothesen, Reflexion sowie die damit verbundene Betrachtung auf der Mikroebene werden – anders als bei der Skizze nach Cooper (2017) – beim *wissenschaftlichen Skizzieren* ausgeschlossen. Die Objektivität von *wissenschaftlichen Skizzen* hängt von der Wahrnehmung ab und wird durch die Fähigkeit, objektiv zu beobachten, beeinflusst.

Wissenschaftliches Skizzieren ist definiert als die Fähigkeit, während des Experimentierprozesses objektive Skizzen anzufertigen. Dadurch, dass eine *wissenschaftliche Skizze* die wesentlichen Aspekte, beispielsweise eines Phänomens, darstellt (Cooper et al., 2017; Kozma & Russell, 2005; Quillin & Thomas, 2015), bietet sie eine förderliche Basis für die Erkenntnisgewinnung (siehe Kapitel 3.2). Gute *wissenschaftliche Skizzen* können fast alle Teilprozesse im Sinne des Teilprozessansatzes (Gut & Mayer, 2018) des Experimentierens, die mit den Teilschritten des wissenschaftlichen Denkens gleichgesetzt werden können, dokumentieren und so die Erkenntnisgewinnung der experimentellen Handlung unterstützen. Ausgeschlossen werden jedoch im Folgenden die Teilprozesse „Generieren von Hypothesen“ und „Reflexion“. Im Teilprozess „Hypothesen generieren“ greifen Lernende auf Vorwissen zurück und entwickeln daraus eine Vermutung und eine Fragestellung (Cooper et al., 2017; Roth & McGinn, 1998). Im Teilprozess „Reflektieren“ verknüpfen die Lernenden Vorwissen und Fachwissen mit Betrachtetem oder greifen die Betrachtung auf der Mikroebene auf. Dadurch zeigen sowohl die grafische Darstellung und Dokumentation als auch die Skizzen dieser beiden Teilprozesse Abhängigkeiten von Vor- und Fachwissen.

Ausgehend zum einen von den theoretischen Grundlagen (siehe Kapitel 3) zur experimentellen Kompetenz, Erkenntnisgewinnung, zu den Methoden der Erkenntnisgewinnung im Allgemeinen, zur Repräsentationskompetenz RC wie auch zu den verschiedenen Funktionen des Skizzierens in schulischen und universitären Kontexten des naturwissenschaftlichen Unterrichts und zum anderen von der hier verwendeten Definition des *wissenschaftlichen Skizzierens*, werden in diesem Kapitel Überlegungen zur Modellierung der Kompetenz des *wissenschaftlichen Skizzierens* dargelegt.

Nachfolgend wird die Kompetenz des *wissenschaftlichen Skizzierens* in den unterschiedlichen Betrachtungsweisen, die in den theoretischen Grundlagen erläutert wurden, modelliert und abgegrenzt: Diese sind die Erkenntnisgewinnung (siehe Kapitel 3.2), die experimentelle Kompetenz (siehe Kapitel 3.1.2), die Repräsentationskompetenz (siehe Kapitel 3.3) sowie externe Faktoren, wie die Kognition.

Tab. 3 zeigt die Vergleichsdimension „Verhältnis zum *wissenschaftlichen Skizzieren*“ für die verschiedenen Kompetenzen, bezüglich derer das *wissenschaftliche Skizzieren* nachfolgend modelliert und abgegrenzt wird. Das Verhältnis mit der experimentellen Kompetenz kann aus zwei unterschiedlichen Blickwinkeln betrachtet werden: einmal aus der Perspektive der experimentellen Kompetenz und zum anderen aus der Per-

spektive des *wissenschaftlichen Skizzierens*. Die experimentelle Kompetenz wird im Forschungsdiskurs unterschiedlich breit aufgefasst und besprochen. Auf der einen Seite des Spektrums wird nur die Umsetzung der VKS (siehe Kapitel 3.2.2.3) explizit als experimentelle Kompetenz angesehen. Auf der anderen Seite des Diskurses wird experimentelle Kompetenz offener definiert und beinhaltet, wie Gut und Mayer (2018) das im Teilprozessansatz (siehe Kapitel 3.1.2) diskutieren, auch den Teilschritt der Dokumentation der Durchführung und Beobachtung sowie das praktische Handeln, also beispielsweise das Beobachten und Vergleichen (siehe Kapitel 3.2.2.1 und 3.2.2.2). Für die Vergleichsdimension in Tab. 3 wird von einer offenen Definition der experimentellen Kompetenz ausgegangen.

Tab. 3 Vergleichsdimension (Verhältnis der experimentellen Kompetenz, der Erkenntnisgewinnung und der RC zum wissenschaftlichen Skizzieren)

Erkenntnis- gewinnung (siehe Kapitel 3.2)	förderliche Basis, um Erkenntnisgewinnung zu erreichen: <i>wissenschaftliche Skizzen</i> stellen die korrekte Oberflächenstruktur eines Phänomens oder Sachverhalts dar, woraus Erkenntnis gewonnen werden kann.	
experimentelle Kompetenz (siehe Kapitel 3.1.2)	Blickwinkel: experimentelle Kompetenz	experimentelle Kompetenz kommt gänzlich ohne skizzieren aus, da Dokumentieren auch nur schriftlich geschehen kann.
	Blickwinkel: <i>wissenschaftli- ches Skizzieren</i>	<i>Wissenschaftliches Skizzieren</i> dokumentiert experimentelle Prozesse, daher ist <i>wissenschaftliches Skizzieren</i> nicht ohne die experimentelle Kompetenz möglich. <i>Wissenschaftliches Skizzieren</i> und experimentelle Kompetenz sind psychometrisch voneinander abhängig.
RC (siehe Kapitel 3.3)	<i>Wissenschaftliches Skizzieren</i> ist ein Bestandteil der RC, wobei RC den Kompetenzbereich Kommunikation umfasst, das <i>wissenschaftliche Skizzieren</i> hingegen nicht. Die <i>wissenschaftlichen Skizzen</i> jedoch bieten die Grundlage für Erkenntnisgewinnungsprozesse.	

4.1 Modellierung in Bezug zur Erkenntnisgewinnung

Wissenschaftliche Skizzen stellen eine mögliche Basis für die Erkenntnisgewinnung dar und unterstützen die Erkenntnisgewinnungsprozesse, da durch eine *wissenschaftliche Skizze* Evidenzen dokumentiert werden. Von einem Phänomen (experimentelle Handlung, z.B. Beobachtung) wird die korrekte Oberflächenstruktur dargestellt und dabei das Wesentliche vom Unwesentlichen unterschieden. Ausgehend von den Qualitätsmerkmalen für eine gute *wissenschaftliche Skizze* (siehe Kapitel 3.5.2), welche

die Voraussetzung bilden, dass aus einer *wissenschaftlichen Skizze* Erkenntnis gewonnen werden kann, werden Grundanforderungen an *wissenschaftlichen Skizzen* abgeleitet und eine daraus mögliche Erkenntnisgewinnung aufgezeigt (siehe Tab. 4).

Die Grundanforderungen decken die Bereiche Eindeutigkeit, Idealisierung, Vollständigkeit, Abstraktion und fachliche Korrektheit ab (siehe Tab. 4 und 8).

In der naturwissenschaftlichen Praxis findet das *wissenschaftliche Skizzieren* u.a. in den Erkenntnismethoden (siehe Kapitel 3.2.2) Beobachten, Vergleichen und Experimentieren statt, daher werden nachfolgend diese drei Erkenntnisgewinnungsmethoden mit den Grundanforderungen an eine *wissenschaftliche Skizze* diskutiert.

Tab. 4 Grundanforderungen und resultierende Erkenntnisgewinnung

Grundanforderung	resultierende Erkenntnisgewinnung
Eindeutigkeit	erkennbar, nachvollziehbar, dass das Phänomen/die Phänomene erzeugt wird/werden bzw. welche Variablen kontrolliert werden
Idealisierung	erkennbar, nachvollziehbar, Relevanz/Bezug zur Fragestellung
Vollständigkeit	Phänomen/Phänomene bzw. VKS erkennbar, nachvollziehbar, alle relevanten Qualitätsmerkmale dargestellt
Abstraktion	erkennbare Abstraktion der relevanten Qualitätsmerkmale
Korrektheit	fachliche Korrektheit der relevanten Qualitätsmerkmale bzw. VKS erkennbar/nachvollziehbar

Beim kriteriengeleiteten Beobachten (siehe Kapitel 3.2.2.1) ist wesentlich, dass die relevanten Beobachungskriterien erkannt und dargestellt werden, damit daraus Erkenntnisse gewonnen werden können. Beim kriteriengeleiteten Vergleichen (siehe Kapitel 3.2.2.2) ist zusätzlich essenziell, dass die Kriterien für den Vergleich wie auch die Unterschiede der beiden Phänomene dargestellt werden. Beim Experimentieren (siehe Kapitel 3.2.2.3) sind wesentliche Kriterien, welche Variablen kontrolliert werden, und ebenso, wie die Variablenkontrolle umgesetzt wird. Gute *wissenschaftliche Skizzen* sind eine förderliche Basis für die Erkenntnisgewinnung, da sie die relevanten Aspekte, auf deren Basis Erkenntnis gewonnen werden kann, abbilden. Das skizzierende Festhalten von Zusammenhängen, wie dies beim Experimentieren (siehe Kapitel 3.2.2.3) der Fall ist, kann bereits ein Teilschritt der Erkenntnisgewinnung sein, welche dadurch dokumentiert wird.

4.2 Modellierung in Bezug zur experimentellen Kompetenz

Experimentelle Handlungen haben eine Erkenntnisgewinnung zum Ziel: Diese kann durch adäquate Durchführung der experimentellen Handlung erreicht werden. Eine gute *wissenschaftliche Skizze* unterstützt die Erkenntnisgewinnung, da sie die relevanten Aspekte, auf deren Basis Erkenntnis gewonnen wird abbildet.

Die Modellierung des *wissenschaftlichen Skizzierens* in Bezug zur experimentellen Kompetenz wird nachfolgend nur aus dem Blickwinkel des *wissenschaftlichen Skizzierens* näher betrachtet (siehe Tab. 3).

Wissenschaftliches Skizzieren und Experimentieren bzw. praktisches Handeln sind psychometrisch interdependent da die Kompetenz des *wissenschaftlichen Skizzierens* nicht getrennt von experimentellen Vorgängen und experimentellen Fähigkeiten, wie bspw. der Beobachtungskompetenz, gemessen werden kann. Die Kompetenz des *wissenschaftlichen Skizzierens* kann in einem experimentellen Setting nie eindeutig von experimentellen Kompetenzen getrennt werden. Eine unzulängliche Kompetenz in einem Teilprozess oder über alle Teilprozesse des Experimentierens bzw. praktischen Handelns hinweg schlägt sich in einer unzulänglichen Performanz beim Skizzieren nieder. Versagen Lernende im Beobachten, gelingt es ihnen auch nicht, einen Vorgang zeichnerisch darzustellen. Wer bspw. nicht fähig ist, eine Durchführung korrekt umzusetzen, könnte zwar eine falsche Durchführung unter Umständen korrekt aufzeichnen, jedoch nicht die fachlich korrekte Durchführung zeichnerisch festhalten. Lernende, die nicht über ausreichende Methodenkompetenzen, wie die Fähigkeit zu beobachten oder zu vergleichen, verfügen, werden auch bei Tests ihre Kompetenzen im *wissenschaftlichen Skizzieren* nur unzulänglich zeigen können. Anders könnte es beim Experimentieren im Sinne einer VKS (siehe Kapitel 3.2.2.3) sein. Schülerinnen und Schüler könnten, trotz mangelhafter Kenntnisse der VKS, in der Lage sein, eine korrekte *wissenschaftliche Skizze* einer Planung eines Experiments zu erstellen.

Um erfolgreich die Kompetenz des *wissenschaftlichen Skizzierens* zu nutzen, müssen auch die experimentellen Kompetenzen, insbesondere in einem experimentellen Setting, beherrscht werden. Um die Unabhängigkeit von Kompetenzen des Hands-On Experimentierens zu gewährleisten, müsste auf das Hands-On-Experimentieren verzichtet und stattdessen bspw. mit Video-Vignetten gearbeitet werden.

Ähnlich wie Lesen und Schreiben könnte das *wissenschaftliche Skizzieren* eine übergeordnete Schlüsselkompetenz darstellen, die zusätzlich zur experimentellen Kompetenz während des ganzen Prozesses erforderlich ist und angewandt wird. Je nachdem, welche Zelle der Matrix (siehe Tab. 2) durch die *wissenschaftliche Skizze* festgehalten und dargestellt wird, unterscheiden sich die Anforderungen an die Lernenden: Eine *wissenschaftliche Skizze* einer Beobachtung unterscheidet sich in der Ausprägung der Grundanforderungen von einer *wissenschaftlichen Skizze* eines Experiments (VKS). Bei der zeichnerischen Darstellung von Beobachtungen und Vergleichen ist mehr Interpretationsspielraum bezüglich des Fokus der *wissenschaftlichen Skizze* vorhanden. Die Kompetenzausprägung des Erstellens solcher *wissenschaftlichen Skizzen* könnte man als sprachlastiger bezeichnen, da diese Skizzen Abläufe von Phänomenen darstellen, die auch sprachlich beschrieben werden könnten. Die Kompetenzausprägung des *wissenschaftlichen Skizzierens* einer Planung einer Variablenkontrollstrategie zeigt die Abhängigkeit zu dieser Strategie. Die unterschiedlichen Ausprägungen der Kompetenz des *wissenschaftlichen Skizzierens* könnten unterschiedliche Zusammenhänge mit externen Variablen wie der Kognition mit sich führen (siehe Kapitel 4.4). Den Unterschieden der verschiedenen Ansätze bzw. Kompetenzbereiche des *wissenschaftlichen Skizzierens* wird im Progressionsmodell (siehe Kapitel 6.2) mittels Ausdifferenzierung der Grundanforderungen Rechnung getragen. *Wissenschaftliches Skizzieren* lässt sich in verschiedenen Kompetenzbereichen darstellen, wodurch das *wissenschaftliche Skizzieren* nicht eine übergeordnete Schlüssel-Kompetenz der experimentellen Kompetenz ist.

Zusammenfassend kann gesagt werden: *Wissenschaftliches Skizzieren* dokumentiert experimentelle Handlungen und bildet die Basis für die Erkenntnisgewinnung (siehe Tab. 3), indem Wesentliches wie auch die korrekte Oberflächenstruktur festgehalten wird. *Wissenschaftliches Skizzieren* ist keine Teilkompetenz der experimentellen Kompetenz, sondern eine eigenständige Kompetenz, welche die Erkenntnisgewinnung in experimentellen Settings unterstützt.

4.3 Modellierung in Bezug zur Repräsentationskompetenz (RC)

Durch normative Verordnung im deutschen Kompetenzbereich Kommunikation wird das *wissenschaftliche Skizzieren* als Teilkompetenz der Kommunikation im Sinne der RC (siehe Kapitel 3.3) aufgeführt.

Das *wissenschaftliche Skizzieren*, wie es hier empirisch verwendet wird, deckt jedoch nur die Fähigkeit der Erstellung von Skizzen der RC ab. Skizzen im Sinne der Kommunikation und des Austauschs werden nicht berücksichtigt. Dadurch ist die RC gegenüber dem *wissenschaftlichen Skizzieren* umfangreicher definiert. Die RC schließt den Kompetenzbereich Kommunikation der deutschen Bildungsstandards (Kultusministerkonferenz, 2005a, 2005b, 2005c) ein. Aus dem Blickwinkel der RC liegt der Fokus des Erstellens von Skizzen nicht auf der *Wissenschaftlichkeit der Skizzen* als förderliche Basis für Erkenntnisgewinnungsmethoden, sondern wird genutzt zur Darstellung von chemischen Phänomenen, zur Diskussion von chemischen Problemen (Kozma & Russell, 2005), zum Aufbau von Fachwissen (wie bspw. submikroskopische Zeichnungen) (Stieff & McCombs, 2006; Stieff & Wilensky, 2003), zur Aufdeckung von Fehlkonzepten (Chi et al., 1981) und zur Aufzeichnung von Problemlöseprozessen (Cardella et al., 2006). Bei der RC wird das Skizzieren im Sinne von Vermittlungskompetenzen, also Kommunikationskompetenzen, aber auch als Methode, um andere Kompetenzen zu fördern, verwendet (siehe auch Kapitel 3.3). Zusammengefasst kann man sagen, die RC hat eine kommunikative, epistemische Funktion, während das *wissenschaftliche Skizzieren* eine epistemologische Funktion aufweist.

Viele der Qualitätskriterien (Grundanforderungen) (siehe Kapitel 3.5.2) *wissenschaftlicher Skizzen* basieren auf Kriterien im Umgang mit Repräsentationen. Kompetenzmodelle, wie das von Chang (2018) (siehe Kapitel 3.3), beinhalten zwar das *wissenschaftliche Skizzieren*, decken aber ein markant größeres Feld ab, nämlich die allgemeine Repräsentationskompetenz, bei welcher der Umgang mit Skizzen eine Teilkompetenz darstellt. Kompetenzbereiche des *wissenschaftlichen Skizzierens* werden durch einen Teilaspekt der Repräsentationskompetenz abgedeckt, da die RC die Kompetenzen der Erstellung von neuen Repräsentationen und Skizzen umfasst. Die Aspekte, die als förderliche Basis der Erkenntnisgewinnung dienen, nämlich, das Wesentliche vom Unwesentlichen zu unterscheiden, sowie die korrekte Oberflächenstruktur abzubilden, sind nicht durch die RC abgedeckt.

4.4 Modellierung in Bezug auf externe Faktoren

Wie in Kapitel 4.2 dargelegt, kann die Kompetenz des *wissenschaftlichen Skizzierens* unterschiedliche Ausprägungen haben, je nachdem, welche Zelle der Matrix der naturwissenschaftlichen Arbeitsweisen und des wissenschaftlichen Denkens nach Neh-

ring et al. (2016) (siehe Tab. 2) fokussiert wird *wissenschaftliches Skizzieren* kann je nach naturwissenschaftlicher Arbeitsweise (Beobachten, Vergleichen, Experimentieren) und wissenschaftlichem Denken (Teilprozesse) unterschiedliche Anforderungen an die Lernenden stellen (siehe auch Kapitel 4.2) und dadurch auch unterschiedliche Abhängigkeiten des *wissenschaftlichen Skizzierens* zu den externen Faktoren aufweisen. Die unterschiedlichen Zusammenhänge mit externen Variablen werden nachfolgend für die verschiedenen Ausprägungen der Kompetenz des *wissenschaftlichen Skizzierens* diskutiert.

Es wird auf mögliche Abhängigkeiten des *wissenschaftlichen Skizzierens* von Fachwissen, Kognition, der naturwissenschaftlichen Arbeitsweise wie auch der Lesefähigkeit eingegangen. Dazu werden Zusammenhänge zwischen den externen Faktoren aufgezeigt und anschließend mit der Kompetenz des *wissenschaftlichen Skizzierens* verknüpft.

Wellnitz und Mayer (2016) betonen die Wichtigkeit der systematischen Beobachtung, da jede Prozessphase auf Sinneserfahrungen und damit auf selektiver Wahrnehmung beruht. Es werden nämlich zur objektiven Beschreibung von Daten Beobachtungen herbeigezogen - dies sind z.B. die beobachtbaren Endzusätze oder die auch die Sinnesdaten („was habe ich gesehen“), die systematisch erfasst werden. Wesentlich ist, dass zwischen beobachtungsrelevanten und unwesentlichen Merkmalen unterschieden wird, also eine Selektion vorgenommen wird. Das systematische Beobachten hat beim kategoriengeleiteten Vergleichen dieselbe Relevanz, da ohne Beobachtung keine Daten gewonnen werden können. Zentral ist wiederum, dass zwischen beobachtungsrelevanten und unwesentlichen Merkmalen unterschieden werden kann. Wie in Kapitel 3.2.2.1 ausgeführt, kann das Beobachten durch Präkonzepte und Vorwissen, wie auch Fachwissen, beeinflusst werden, wodurch es abhängig von der Wahrnehmung ist. Die Beeinflussung von Präkonzepten kann durch einen gerichteten Fokus, wie Beobachtungskriterien, abgeschwächt werden, wodurch eine gewisse Objektivität der Beobachtung erreicht werden kann.

Nehring und Schwichow (2020) konnten aufzeigen, dass zwischen den Kompetenzbereichen Fachwissen und prozessorientierten Kompetenzen der Erkenntnisgewinnung ein nicht zu vernachlässigender Zusammenhang besteht. Für die Fachkontexte Biologie, Chemie und Physik geht höhere Kompetenzausprägung der VKS mit höherem Fachwissen und höherer Jahrgangsstufe einher (Schwichow & Nehring, 2018).

Jedoch gehen Schwichow und Nehring (2018) von einer indirekt kausalen Abhängigkeit aus (siehe Kapitel 3.2.2.3), die durch allgemeine kognitive Fähigkeiten bedingt ist.

Zwischen einem Test zur Erfassung von Strategiewissen zum Experimentieren (NAW) (Klos et al., 2008; Koenen, 2014; Mannel, 2011) und dem Fachwissen in Chemie besteht eine sehr geringe Korrelation, daher besteht kein erwähnenswerter Zusammenhang zwischen den erhobenen Konstrukten des NAW (Strategiewissen beim Experimentieren) und dem eingesetzten Fachwissenstest (Klos et al., 2008). Die gemessene, signifikante Korrelation von 0.360 zwischen der allgemeinen Kognition, gemessen mit dem KFT (Heller & Perleth, 2000), und den Leistungen im NAW weisen auf einen Einfluss der allgemeinen Kognition hin (Klos et al., 2008). Ein Zusammenhang zwischen NAW und der verbalen Skala des KFT konnte nicht gezeigt werden (Klos et al., 2008). Der NAW misst weder hauptsächlich kognitive Fähigkeiten noch reines Fachwissen und hängt nicht von der Lesefähigkeit ab, sondern erfasst ein darüber hinaus vorhandenes Konstrukt, wie Lernende in den Naturwissenschaften Denken und Handeln.

Höner und Werzel (2018) konnten einen schwach positiven signifikanten Zusammenhang zwischen den erreichten Punkten in der naturwissenschaftlichen Aufgabensequenz (praktisches Handeln) und den erreichten Punkten in den kognitiven Subskalen des KFT (Heller & Perleth, 2000) zeigen. Die Korrelation für den verbalen Subtest liegt mit 0.327 etwas höher als für den nonverbalen Subtest (N3) mit 0.228. Damit konnte nur die Tendenz gezeigt werden, dass höhere kognitive Fähigkeiten mit besseren Leistungen beim naturwissenschaftlichen Problemlösen verknüpft sind. Ähnliche Ergebnisse zeigte Nehring (2014): Er erkannte Zusammenhänge zwischen Kompetenzen im Bereich der Erkenntnisgewinnung und der kognitiven Intelligenz (0.36) sowie dem Leseverständnis (0.19). Auch Wirth et al. (2005) zeigten eine positive Korrelation zwischen naturwissenschaftlichem Problemlösen und kognitiven Fähigkeiten. Eine höhere kognitive Aktivierung könnte in Anlehnung an Helme und Schrader (2009) zu einem höheren Zusammenhang mit der Kognition führen.

In der Studie ExKoNawi (Bonetti et al., 2018) zeigt sich zwischen dem Strategiewissen zum Experimentieren und dem praktischen Untersuchen ein hochsignifikanter bedeutender Zusammenhang (0.516). Hingegen fällt der Zusammenhang mit 0.222 zwischen Kognition und praktischem Untersuchen weniger bedeutend aus. Bonetti et al.

(2018) zeigten in der ExKoNawi-Studie durch den hohen Zusammenhang zwischen NAW und praktischem Untersuchen, dass ein ähnliches Konstrukt gemessen wurde, welches das Verständnis der Variablenkontrollstrategie voraussetzt.

Ausgehend einerseits von positiven Korrelationen zwischen Kognition und naturwissenschaftlicher Problemlöse-Fähigkeit und andererseits von der Kognition und dem Strategiewissen zum Experimentieren, kann ein theoretischer Zusammenhang der Kognition mit der Kompetenz des *wissenschaftlichen Skizzierens* vermutet werden.

Ein Zusammenhang zwischen dem *wissenschaftlichen Skizzieren* und dem Fachwissen kann nicht ausgeschlossen werden, jedoch scheinen Fähigkeit der VKS als auch auf die Leistungen im NAW-Test vor allem auf den Zusammenhang mit der Kognition zurückzuschließen zu sein.

Die Einbettung der Abhängigkeit vom Fachwissen ändert sich zudem mit dem Anwendungsbereich des Skizzierens. Wird von der Struktur der Skizzen nach Cooper (2012) ausgegangen, wo bei Reflexions-Skizzen die Mikroebene und auch die Teilchenstruktur betrachtet werden, wird das Fachwissen relevanter. Der Einfluss des Fachwissens scheint bei einer Repräsentationsform der Makroebene deutlich geringer. Einen Zusammenhang zwischen Leseverständnis, Kognition und Erkenntnisgewinnungskompetenzen konnten Wirth et al. (2005) zwar zeigen, jedoch könnte das Leseverständnis beim Erstellen von Skizzen wesentlich weniger relevant sein, da Skizzieren weniger sprachaffin ist. Ein Zusammenhang zwischen dem NAW-Test und dem *wissenschaftlichen Skizzieren* wird auf Grund der Abhängigkeit des NAW-Tests von der Kognition vermutet.

4.5 Fazit Modellierung

Die Kompetenz des *wissenschaftlichen Skizzierens* weist enge Verbindungen zur experimentellen Kompetenz wie auch zur RC auf. Auch zwischen der experimentellen Kompetenz und der RC gibt es Zusammenhänge. Die experimentelle Kompetenz ist verknüpft mit weiteren fachspezifischen Kompetenzen (z.B. Fachwissen), mit allgemeinen kognitiven Fähigkeiten sowie dem Modellieren und dem Umgang mit Diagrammen (Gut & Mayer, 2018). Der Umgang mit Diagrammen, wie auch grundsätzlich jede Dokumentation jeder experimentellen Handlung, ist Bestandteil der RC.

Die Kompetenz des *wissenschaftlichen Skizzierens* stellt ein neues Konstrukt dar, das weder durch die experimentelle Kompetenz noch durch die RC vollständig abgedeckt

ist. Zusammenhänge mit der Kognition und mit dem Strategiewissen zum Experimentieren werden angenommen. Jedoch geht man kaum von Zusammenhängen mit der Lesefähigkeit aus. *Wissenschaftliche Skizzen* bieten eine förderliche Basis für die Erkenntnisgewinnung, da Wesentliches durch eine *wissenschaftliche Skizze* festgehalten wird.

5 Forschungsziel und Forschungsfragen

Wie in den Kapiteln 3.4 und 4 aufgezeigt wurde, ist das *wissenschaftliche Skizzieren* in der naturwissenschaftlichen Didaktik bis anhin wenig systematisch erforscht (Nitz et al., 2014). Ein Großteil der Forschung an Skizzieraktivitäten bezieht sich auf die Rolle des Skizzierens als Unterstützung in Assessments, als Erleichterung von naturwissenschaftlichem Modellieren (Cheng, 2018; Cooper et al., 2015; Harle & Towns, 2013; Kelly et al., 2010) oder als Methode, andere Ziele zu erreichen. *Wissenschaftliches Skizzieren* als förderliche Basis für die Erkenntnisgewinnung (Van Meter & Garner, 2005) dagegen ist bislang kaum erforscht. Die zentrale Bedeutung des *wissenschaftlichen Skizzierens* als förderliche Basis der Erkenntnisgewinnung (Cooper et al., 2012, 2017; Frischknecht-Tobler & Labudde, 2010; Quillin & Thomas, 2015; Roth & McGinn, 1998; Van Meter & Garner, 2005) konnte im Kapitel 3 aufgezeigt werden. Eine präzise Definition des Konstrukts des *wissenschaftlichen Skizzierens* als förderliche Basis der Erkenntnisgewinnung experimentellen Handelns wie auch eine Abgrenzung zu anderen Konstrukten wie der experimentellen Kompetenz oder der Repräsentationskompetenz hingegen fehlt bislang (siehe Kapitel 4).

Derzeit ist kaum erforscht, welche Kompetenzen die Lernenden der Sekundarstufe I im Bereich des *wissenschaftlichen Skizzierens* aufweisen, wie sie unterrichtet werden und ebenso wenig, wie sie gefördert werden können (Nitz et al., 2014; Van Meter & Garner, 2005) (siehe Kapitel 3.4.1). In Bezug auf das *wissenschaftliche Skizzieren* als Schülerkompetenz gibt es bislang in der Literatur kein spezifisches Modell, das die Struktur oder die Stufung einer entsprechenden Kompetenz beschreibt. Ansätze zeigt Hope's Framework (2008) (siehe Kapitel 3.5.1). Die Kompetenzmessung weist hier aber eine Lücke auf. *Wissenschaftliches Skizzieren* lässt sich durch Qualitätsmerkmale (siehe Kapitel 3.5.2 und Tab. 8) messen, welche die Grundlage für die Konstruktion eines Kompetenzstrukturmodells bieten. Daher werden diese Qualitätsmerkmale zur Konstruktion des Kompetenzmodell genutzt.

Wahrscheinliche Zusammenhänge des *wissenschaftlichen Skizzierens* je nach Zelle der Matrix nach Nehring et al. (2016) (siehe Tab. 2) mit der Kognition (Nehring, 2014; Wirth et al., 2005), der naturwissenschaftlichen Arbeitsweise (Bonetti et al., 2018) und mit dem Fach- und Methodenwissen (Klos et al., 2008; Schwichow & Nehring, 2018) konnten auf Grund des Zusammenhangs mit den experimentellen Kompetenzen angenommen werden (siehe Kapitel 4.4).

Auf Grund des geringen Zusammenhangs zwischen Kognition und Lesekompetenz (Schneider et al., 2007) und der Tatsache, dass Skizzieren nicht sprachlastig ist, wird von keinen signifikanten Abhängigkeiten von Lesekompetenzen ausgegangen (siehe Kapitel 4.4).

Das Ziel des Dissertationsprojekts ist es, zur Schließung dieser Lücke in der Forschung beizutragen. Dafür wurde sowohl ein geeignetes Kompetenzmodell als auch, darauf aufbauend, ein geeignetes Diagnoseinstrument entwickelt und evaluiert, mit dem die Fähigkeiten des *wissenschaftlichen Skizzierens* bei Lernenden untersucht werden können. Der Themenbereich wurde dabei auf die Chemie in experimentellen Handlungen im Bereich Sekundarstufe I eingeschränkt.

Das dafür erforderliche Forschungsdesign wurde auf das Experimentieren und experimentelle Handeln reduziert. *Wissenschaftliches Skizzieren* im Sinne von Modellieren und Argumentieren wurde im Forschungsdesign nicht berücksichtigt, da das Modellieren keine klassischen Dokumentieranlässe im Schulfeld darstellt (siehe Kapitel 3.2.1 und 3.4.3).

Tab. 5 Forschungsdesign mit Aufgabentypen

<i>Naturwissenschaftliche Arbeitsweisen</i>			
Experimentieren		Aufgabentyp 3	
Beobachten, Vergleichen, Ordnen		Aufgabentyp 1, Aufgabentyp 2	
	Fragestellung & Hypothesen	Planung & Durchführung	<i>Wissenschaftliches Denken</i>

Ausgehend von der Struktur eines Kompetenzmodells zur fachübergreifenden Vernetzung der Erkenntnisgewinnung (VerE-Modell) (Nehring et al., 2016) (siehe Tab. 2) und den Aufgabenstruktur nach Cooper (2017) und Cavelti (2018) wurde das Forschungsdesign entwickelt: Die naturwissenschaftlichen Arbeitsweisen Beobachten (beobachtende Skizze nach Cooper et al. (2017)) und Experimentieren (voraussagende bzw. Planungs-Skizze nach Cooper et al. (2017)) wurden matrixartig mit dem wissenschaftlichen Denken, wie bspw. Planung und Durchführung/Beobachtung, verknüpft. Fokussiert wurden exemplarisch drei Zellen dieser Matrix, woraus drei Aufgabentypen resultieren (siehe Tab. 5). Die Zelle Beobachten, Vergleichen vs. Durchführung mit den Aufgabentyp 1 „Beobachtung“ (Beobachtung eines Phänomens) und Aufgabentyp 2 „Vergleich“ (Beobachtung des Vergleichs zweier Phänomene) decken

zudem die Methoden der Erkenntnisgewinnung (Kultusministerkonferenz, 2005b) Beobachtung und Vergleich ab und schließen an die Beobachtungs-zeichnung nach Cooper et al. (2017) und Cavelti et al. (2018) an. Die Zelle Experimentieren vs. Planung mit dem Aufgabentyp 3 „Experimentieren“ deckt das Experimentieren zudem als Erkenntnisgewinnungsmethode ab und entspricht der Planungszeichnung nach Cavelti et al. (2018). Die Kompetenz des *wissenschaftlichen Skizzierens* wird unter Berücksichtigung der unterschiedlich ausgeprägten Grundanforderungen (siehe auch Kapitel 4.4 und Tab. 4) für diese Aufgabentypen im Kompetenzstrukturmodell ausdifferenziert (siehe Kapitel 6.1).

Die exemplarische Auswahl der Zellen wurde gemäß der theoretischen Verordnung und Abgrenzung des *wissenschaftlichen Skizzierens* (siehe Kapitel 4) vorgenommen. Die Aufgabentypen Beobachtung“ und „Vergleich“ decken die gleichnamigen Methoden der Erkenntnisgewinnung ab und schließen an die Problemtypen kategoriengeleitetes Beobachten an (Gott & Welford, 1987; Hammann, 2002; Solano-Flores et al., 1997; Stevens, 1978) (siehe Kapitel 3.2.2.1 und 3.2.2.2), welche mit den Teilaufgaben „ein einzelnes Phänomen beschreiben“ und „Unterschiede und Gemeinsamkeiten zweier Phänomene beschreiben“ beschrieben werden (Metzger & Gut, 2017). Aus dem Blickwinkel der naturwissenschaftlichen Arbeitsweisen und des wissenschaftlichen Denkens (siehe Kapitel 3.1.2.) werden die Beobachtung und der Vergleich als naturwissenschaftliche Arbeitsweisen und die Durchführung als wissenschaftliches Denken in diesen beiden Aufgabentypen abgedeckt. Der Aufgabentyp „Experimentieren“ nimmt das Experimentieren als Erkenntnisgewinnungsmethode und den Problemtyp (Gut & Mayer, 2018) der experimentellen Kompetenz „fragengeleitetes Untersuchen“ (siehe Kapitel 3.2.2.3) auf. Die naturwissenschaftlichen Arbeitsweisen des Experimentierens und die Planung des wissenschaftlichen Denkens werden in diesem Aufgabentyp berücksichtigt.

Ausdifferenziert werden die drei Aufgabentypen in je drei Kontexten (Brausetablette, Eis schmelzen und Säuren und Basen) (siehe auch Kapitel 6.4.3), welche die curriculare Verordnung des Lehrplan 21 (D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz, 2016b) entsprechen Zur Untersuchung der Abgrenzung von externen Faktoren (siehe Kapitel 4.4) wurden begleitend die Faktoren Kognition, Strategiewissen zum Experimentieren und Lesefähigkeit erfasst. Die Messung der Kompetenz des

wissenschaftlichen Skizzierens erfolgte in einem Hands-on-Setting mit Video-Vignetten (siehe Kapitel 6.3.2).

Basis zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* ist ein Kompetenzstrukturmodell (siehe Kapitel 3.5 und Kapitel 6.2), welches als Grundlage die Qualitätsmerkmale (siehe Kapitel 3.5.2) *wissenschaftlicher Skizzen*, die Definition (siehe Kapitel 3.4 und 4) wie auch Abgrenzung zu anderen Kompetenzen wie der experimentellen Kompetenz (siehe Kapitel 3.1.2) und der RC (siehe Kapitel 3.3) aufweist.

Die Fragestellungen lassen sich in die Bereiche 1) Entwicklung des Messinstruments, 2) Überprüfung von Testgüte, Messinvarianz und Generalisierbarkeit des Messinstruments sowie 3) externe Validierung des Messinstruments unterteilen.

1) Entwicklung des Messinstruments:

Für die Entwicklung des Messinstruments ergeben sich die folgenden beiden Forschungsfragen:

F1: Sind die Qualitätsmerkmale des entwickelten Kompetenzstrukturmodells inhaltlich valide?

F2: Sind die für die Testkonstruktion verwendeten Kontexte und Aufgaben ökologisch valide?

2) Überprüfung von Testgüte, Messinvarianz und Generalisierbarkeit des Messinstruments:

Unter der Voraussetzung, dass Qualitätsmerkmale des entwickelten Kompetenzstrukturmodells inhaltlich und die für die Testkonstruktion verwendeten Kontexte und Aufgaben ökologisch valide sind, kann das neu entwickelte Messinstrument hinsichtlich Testgüte, Messinvarianz und Generalisierbarkeit überprüft werden. Dieser Überprüfung wird in der Forschungsfrage 3 Rechnung getragen.

F3: Erfüllt das Messinstrument die Anforderungen hinsichtlich Testgüte, Messinvarianz und Generalisierbarkeit?

3) Externe Validierung des Messinstruments:

Unter der Voraussetzung, dass das neu entwickelte Messinstrument die Bedingungen der Testgüte und Messinvarianz erfüllt und Generalisierbarkeit zulässt, wird die externe Validität des Messinstruments überprüft. In Anlehnung an die Resultate des Pro-

jekt ExKoNawi (Bonetti et al., 2018) (siehe Kapitel 4.4) und an die Studie von Schwichow und Nehring (2018) (siehe Kapitel 3.2.2.3) lässt sich vermuten, dass Kognition und Strategiewissen zum Experimentieren (VKS), die Sprachfähigkeit jedoch nicht, einen Zusammenhang mit der Kompetenz des *wissenschaftlichen Skizzierens* haben (siehe auch Kapitel 4.1). Ausgehend von einem Diagnoseinstrument, das in Bezug auf Testgüte, Messinvarianz und Generalisierbarkeit genügt, kann die vierte Forschungsfrage bezüglich der externen Validität abgeleitet werden.

F4: Besteht ein Zusammenhang zwischen dem *wissenschaftlichen Skizzieren* und der Kognition (KFT), dem Strategiewissen zum Experimentieren (NAW) und der Lesefähigkeit (LGVT)?

Unter Berücksichtigung sowohl des Projekts ExKoNawi (Bonetti et al., 2018) (siehe Kapitel 4.4), das zeigte, dass Kognition und Strategiewissen zum Experimenten auf die Performance in Hands-on Experimenten einen positiven Zusammenhang aufweisen (siehe Kapitel 4.4), als auch der Forschungsergebnisse von Nehring (2014) und Wirth et al. (2005), die einen positiven Zusammenhang von Kognition mit der Erkenntnisgewinnung und ebenso eine Abhängigkeit des Strategiewissens von der Kognition zeigen konnten (siehe Kapitel 4.4), werden in Hinblick auf die Forschungsfrage die vier folgenden Hypothesen aufgestellt:

H4.1: Die Kognition steht in einem mittleren positiven Zusammenhang mit *wissenschaftlichem Skizzieren* bei allen drei Aufgabentypen.

H4.2: Das Strategiewissen zum Experimentieren steht in einem schwachen positiven Zusammenhang mit *wissenschaftlichem Skizzieren* bei allen drei Aufgabentypen.

H4.3: Die Sprachfähigkeit steht in keinem Zusammenhang mit *wissenschaftlichem Skizzieren* bei allen drei Aufgabentypen, da skizzieren kaum sprachliche Anforderungen an die Lernenden stellt. Außerdem wurde der Aufgabenstamm sprachlich sehr einfach formuliert.

Dem Zusammenhang des Multiple-Select Test, welcher das Fachverständnis der Kontexte – also das Beobachten und das Beobachten von Vergleichen der Aufgabentypen „Beobachtung“ und „Vergleich“ und das Methodenwissen des Aufgabentyps „Experimentieren“ – aufgreift, mit der Kompetenz des *wissenschaftlichen Skizzierens*, wird in der Forschungsfrage 5 nachgegangen.

F5: Besteht ein Zusammenhang zwischen dem relevanten Fach- und Methodenwissen und der Kompetenz des *wissenschaftlichen Skizzierens*?

Im Hinblick auf die Forschungsfrage 5 werden im Folgenden zwei Hypothesen aufgestellt. Dabei wird zum einen berücksichtigt, dass die Beobachtungskompetenz und somit das Fachwissen, das mit dem Multiple-Select Test (siehe Kapitel 6.3.4) abgefragt wird, mit der Qualität der Skizzen der Aufgabentypen „Beobachtung“ und „Vergleich“ korreliert und zum anderen vermutet, dass eine indirekte kausale Abhängigkeit zwischen dem Methodenwissen der VKS und dem Fachwissen, das vor allem auf die Kognition zurückzuführen ist (Schwichow & Nehring, 2018) (siehe Kapitel 3.2.2.3), besteht.

H5.1: Das relevante Fach- und Methodenwissen der Aufgabentypen „Beobachtung“ und „Vergleich“ steht in einem mittleren positiven Zusammenhang mit *wissenschaftlichem Skizzieren* der Aufgabentypen „Beobachtung“ und „Vergleich“.

H5.2: Das relevante Fach- und Methodenwissen des Aufgabentyps „Experimentieren“ steht in einem, wenn überhaupt, schwachen Zusammenhang mit *wissenschaftlichem Skizzieren* des Aufgabentyps „Experimentieren“, da vor allem bei jüngeren Lernenden eine ausgeprägte Methodenkompetenz der Planung eines Experiments noch nicht vorhanden ist (Schwichow & Nehring, 2018; Sodian & Mayer, 2013) (siehe Kapitel 3.2.2.3).

Im Projekt ExKoNawi (Bonetti et al., 2018) konnte gezeigt werden, dass Problemtypen (Gut & Mayer, 2018) unterscheidbar sind: Daraus leitet sich die Forschungsfrage 6 ab.

F6: Besteht ein Unterschied zwischen den Zusammenhängen der verschiedenen Aufgabentypen des *wissenschaftlichen Skizzierens* mit der Kognition (KFT) und dem Strategiewissen zum Experimentieren (NAW)?

Im Hinblick auf die Forschungsfrage 6 und ausgehend von den im Kapitel 4 vermuteten und beschriebenen Kompetenzbereichen des *wissenschaftlichen Skizzierens* wird folgende Hypothese aufgestellt:

H6.1: Es besteht ein Unterschied bezüglich der Zusammenhänge von Kognition (KFT) und Strategiewissen zum Experimentieren (NAW) mit dem *wissenschaftlichen Skizzieren* bei den Aufgabentypen „Beobachtung“ und „Vergleich“ im Vergleich zum Aufgabentyp „Experimentieren“ (Gut et al., 2014).

6 Methoden

Basierend auf der theoretischen Verordnung des *wissenschaftlichen Skizzierens* (siehe Kapitel 4) wurde zur Erreichung der in Kapitel 5 beschriebenen Forschungsfragen F1 - F6 ein Kompetenztest (Video-Vignette und Paper-Pencil-Test) eingesetzt, der die Kompetenz des *wissenschaftlichen Skizzierens* misst. Es wurde ein Aufgabenset entwickelt, das die Kompetenzbereiche des *wissenschaftlichen Skizzierens* in den Aufgabentypen „Beobachtung“, „Vergleich“ und „Experimentieren“ (siehe Kapitel 5) erfasst. Ein Item umfasst eine Video-Vignette mit einem Skizzierauftrag und einen neu entwickelten Multiple-Select Test zur Erfassung des Fach- und Methodenwissens (siehe Kapitel 6.3.4) im Kontext des jeweiligen Items.

Zur Beantwortung der inhaltlichen wie ökologischen Validität (siehe Forschungsfragen F1 und F2) wurden Expertenbefragungen bzw. Befragungen von Lehrpersonen (siehe Kapitel 7.2 und 7.3) durchgeführt.

Die Itemschwierigkeiten und die Modellpassung (siehe F3) des neu konstruierten Tests wurden mittels einer eindimensionalen Rasch-Analyse überprüft (siehe Kapitel 6.8). Dieser Ansatz entspricht der Modellierung durch die Qualität der Anforderungsbewältigung. Die Items des neu konstruierten Tests wurden derart generiert, dass sie möglichst gleichwertige Anforderungen an die Lernenden stellen. Die Validierung ist die Bewertung, wie gut die Items gelöst werden. Die Validierungsaspekte Generalisierbarkeit, insbesondere die Invarianz des Aufgabensets (siehe F3), wurden durch differenzierte Rasch-Analysen (DIF) überprüft (siehe Kapitel 6.8 und 7.4).

Bei der Entwicklung des Messinstruments wurden die drei Gütekriterien Objektivität, Reliabilität und Validität berücksichtigt (siehe Kapitel 6.7). Die Objektivität beinhaltet die Durchführungsobjektivität und die Auswertungsobjektivität. Der Durchführungsobjektivität wurde durch die Unabhängigkeit der Rahmenbedingungen (z.B. personenunabhängige Durchführung) Rechnung getragen. Auch die Unabhängigkeit von konfundierenden Variablen, wie Lernumgebung oder Motivation der Lernenden, wurde, soweit wie beeinflussbar, berücksichtigt, indem immer dasselbe standardisierte Testverfahren gewählt wurde. Die Auswertungsobjektivität wurde unter dem Aspekt der Interrater-Reliabilität Rechnung getragen (siehe Kapitel 6.6).

Basierend auf einem Messinstrument, das die Anforderungen in Bezug auf Messinvarianz und Generalisierbarkeit erfüllt, wurde die externe Validität untersucht (siehe

Kapitel 6.7). Zur Überprüfung der im Kapitel 5 beschriebenen Hypothesen zur externen Validität wurden folgende Tests eingesetzt:

- Kompetenztest (Video-Vignette und Paper-Pencil-Test)
- Multiple-Select Test zur Erfassung des Fach- und Methodenwissens
- Lesegeschwindigkeits-Leseverständnistext (LGVT) (Schneider et al., 2007)
- Test zur Erfassung des Strategiewissen zum Experimentieren (NAW) (Koenen, 2014; Mannel, 2011)

sowie ein Fragebogen zur Erfassung der kognitiven Faktoren (KFT) (Heller & Perleth, 2000) .

Die Beantwortung der F4 erfolgte mit Regressions- und Strukturgleichungsanalysen der externen Faktoren (siehe Kapitel 6.9 - 6.11). Zur Beantwortung der F5 wurde der Zusammenhang mit den begleitend erhobenen Fach- und Methodentests (siehe Kapitel 6.10) untersucht, und für die F6 wurden mögliche Unterschiede der Zusammenhänge zu den externen Variablen mittels einer zweidimensionalen Rasch- und einer Strukturgleichungsanalyse untersucht (siehe Kapitel 6.11). Im Folgenden werden die eingesetzten Instrumente vorgestellt, deren Konstruktion erläutert und die Auswertung der eingesetzten Instrumente erklärt.

6.1 Entwicklung des Kompetenzstrukturmodells

Um die Kompetenz des *wissenschaftlichen Skizzierens* (siehe Kapitel 4) der Schülerinnen und Schüler der Sekundarstufe I zu erfassen, bedarf es der Entwicklung standardisierter Testaufgaben auf der Grundlage eines Kompetenzstrukturmodells (siehe Kapitel 3.5). Ein Kompetenzstrukturmodell ermöglicht es, Kompetenzausprägungen zu einem bestimmten Zeitpunkt zu beschreiben und die Kompetenzstruktur der einzelnen Kompetenzstufen darzustellen. Außerdem kann abgebildet werden, welche Kompetenzen bei den verschiedenen Niveaus erreicht werden können. Im Folgenden wird erst die Kompetenzstruktur der Aufgabentypen und die innere Differenzierung Kompetenzstruktur vorgestellt und darauf aufbauend die aufgabentypspezifische Progression erläutert.

6.1.1 Kompetenzstruktur der Aufgabentypen

Die Kompetenz *wissenschaftlichen Skizzierens* wurde für die Schnittmengen der Matrix der naturwissenschaftlichen Arbeitsweisen (Nehring et al., 2016) bzw. der Pro-

blemtypen (Gut & Mayer, 2018) und des wissenschaftlichen Denkens (Nehring et al., 2016) bzw. der Teilprozesse (Gut & Mayer, 2018) des experimentellen Handelns (siehe Kapitel 3.1.2 und Kapitel 5) ausdifferenziert. Zusammen mit den unterschiedlichen Kontexten bilden sie die Struktur des Kompetenzstrukturmodells. Jeder Aufgabentyp fokussiert sowohl einen Teilprozess (oder wissenschaftliches Denken) im Experimentierablauf als auch eine naturwissenschaftliche Arbeitsweise (oder einen Problemtypen (Gut & Mayer, 2018)) und wurde in Anlehnung an die Aufgabenstruktur von Cooper et al. (2017) und die Struktur des Dokumentier-Führerscheins (Cavelti et al., 2018) entwickelt (siehe Kapitel 5). Davon ausgehend, dass die Kompetenz des *wissenschaftlichen Skizzierens* (siehe Kapitel 4), je nach ausgewähltem Aspekt (Schnittmenge aus naturwissenschaftlicher Arbeitsweise und wissenschaftlichem Denken), der im Fokus liegt, wahrscheinlich unterschiedlich sein könnte, wurden drei Aufgabentypen gewählt.

Die ersten beiden Aufgabentypen „Beobachtung eines Phänomens“ (nachfolgend „Beobachtung“ genannt) und „Vergleich zweier Phänomene“ (nachfolgend „Vergleich“ genannt) fokussieren den Teilprozess der Durchführung einer experimentellen Handlung wie auch den Problemtyp „kategoriengeleitetes Beobachten“ und entsprechen der beobachtenden Skizze der Aufgabenstruktur von Cooper et al. (2017) (siehe Kapitel 5). Beim Aufgabentyp „Beobachtung“ gilt es, das erzeugte Phänomen in einer *wissenschaftlichen Skizze* festzuhalten (siehe auch 3.2.2.1). Beim Aufgabentyp „Vergleich“ sollen beiden erzeugten Phänomene sowohl in einer *wissenschaftlichen Skizze* festgehalten und mit der Skizze ein Vergleich dargestellt werden (siehe auch 3.2.2.1 und 3.2.2.2). Wie bereits beim Dokumentier-Führerschein (Cavelti et al., 2018) wurde bei der voraussagenden Skizze (Cooper et al., 2017) nur der Planungsaspekt für eine fragengeleitete Untersuchung (Gut & Mayer, 2018) berücksichtigt, da die Erhebung von Präkonzepten oder Fachwissen vermieden werden soll.

Der dritte Aufgabentyp „Planung eines Experiments“ (nachfolgend „Experimentieren“ genannt) greift sowohl den Teilprozess „Planung eines Experiments“ und den Problemtyp „fragengeleitetes Untersuchen“ (Gut et al., 2014; Gut & Mayer, 2018) als auch die Planung von Untersuchungen als Teilkompetenz (Vorholzer et al., 2016) auf (siehe Kapitel 5). Beim Aufgabentyp „Experimentieren“ liegt der Fokus auf der zeichnerischen Darstellung der Experimentieranordnung und des Experimentierablaufs,

wobei es sich beim Experiment um die Untersuchung eines Zusammenhangs handelt und dabei die VKS aufgegriffen wird (siehe auch 3.2.2.3).

6.1.2 Innere Differenzierung der Kompetenzstruktur

Die Grundanforderungen der Kompetenz des *wissenschaftlichen Skizzierens* (siehe Tab. 4) können Kategorien zugeordnet werden. Alle Aufgabentypen finden sich in der Kategorie „*Wissenschaftlichkeit der Skizze*“, die in fünf Facetten aufgeteilt wird. Für die Aufgabentypen „Beobachtung“ und „Vergleich“ sind diese fünf Facetten (siehe Tab. 6) dieser Kategorie inhaltlich identisch. Beim Aufgabentyp „Experimentieren“ liegen aufgrund der Fokussierung auf die Planung leichte Abweichungen in der inhaltlichen Beschreibung der einzelnen Facetten vor.

Tab. 6 zeigt die Kategorien wie auch die Aufgabenstellung der Aufgabentypen auf.

Tab. 6 Kategorien und Aufgabenstellung der Aufgabentypen

Aufgabentyp	Aufgabenstellung	Kategorien	
Beobachtung	erzeugtes Phänomen skizzieren	Wissenschaftlichkeit der Skizze (aufgeteilt in fünf Facetten)	
Vergleich	erzeugte Phänomene vergleichen und skizzieren	Wissenschaftlichkeit der Skizze (aufgeteilt in fünf Facetten)	Unterschied erfasst und Vergleich dargestellt
Experimentieren	Experimentieranordnung und Experimentierablauf skizzieren	Wissenschaftlichkeit der Skizze (aufgeteilt in fünf Facetten)	Variablen erfasst und vollständig kontrolliert

Die beiden Aufgabentypen „Vergleich“ und „Experimentieren“ verfügen im Kompetenzstrukturmodell über je eine zusätzliche Kategorie, welche die aufgabentypspezifischen Qualitätsmerkmale beschreibt. Diese sind die Darstellung des Vergleichs und die Erfassung des Unterschieds für den Aufgabentyp „Vergleich“ und die Erfassung und Kontrolle der beiden Variablen für den Aufgabentyp „Experimentieren“. Diese aufgabentypspezifischen Kategorien ergänzen die allgemeine Kategorie „*Wissenschaftlichkeit der Skizze*“. Die Kategorie „Unterschied erfasst und Vergleich dargestellt“ greift erweiternd die Fähigkeiten im Bereich der zeichnerischen Darstellung des Unterschieds auf. Die Kategorie „Variablen erfasst und vollständig kontrolliert“ erfasst die Fähigkeit, die Variablenkontrollstrategie (siehe Kapitel 3.2.2.3) in einer

wissenschaftlichen Skizze darzustellen, also ob Lernende nur eine einzelne oder beide Variablen erfassen und auch deren vollständige Kontrolle darstellen können.

Tab. 7 Beschreibung der Facetten der Kategorie „Wissenschaftlichkeit der Skizze“

Facette	Beschreibung
Eindeutigkeit	Sind die relevanten Qualitätsmerkmale zum Herstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt?
Idealisierung	Werden nur relevante Objektmerkmale fokussiert?
Vollständigkeit	Sind alle relevanten Objektmerkmale der Reaktion dargestellt?
Abstraktion	Sind die relevanten Objektmerkmale genügend abstrahiert?
Korrektheit	Sind die relevanten Objektmerkmale fachliche korrekt dargestellt?

Die Kategorie „*Wissenschaftlichkeit der Skizze*“ (siehe Tab. 7) beschreibt die Grundanforderungen an das *wissenschaftliche Skizzieren* und wird im Kompetenzstrukturmodell theoriebasiert (siehe Kapitel 6.2) in fünf Facetten unterteilt. Die Tab. 7 zeigt die Beschreibung der Facetten der Kategorie „*Wissenschaftlichkeit der Skizze*“.

Wie erwähnt, verfügt die Kategorie „*Wissenschaftlichkeit der Skizze*“ in allen Aufgabentypen über fünf Facetten, welche die Struktur dieser Kategorie abbilden. Es sind dies Eindeutigkeit, Idealisierung der Skizze, Vollständigkeit der Skizze, Abstraktion der Skizze und Korrektheit der Skizze.

Diese Kategorie wird aus theoretischen Überlegungen auf alle Aufgabentypen gleichermaßen als Set von Qualitätsmerkmalen angewandt, wobei die Ausdifferenzierung in den Kodierschemen (siehe Anhang) jeweils kontextspezifisch (wenn es um das Herstellen eines Phänomens oder den Ablauf eines Phänomens geht) erfolgen muss/erfolgt. Die anderen beiden Kategorien sind nicht nur aufgabentypabhängig, da sie spezifisch für einen Aufgabentypen sind, sondern auch kontextabhängig, da die Unterschiede, beziehungsweise die zu kontrollierenden Variablen, sich von Kontext zu Kontext unterscheiden. Die Kontextabhängigkeit wird jedoch nur im Kodiermanual berücksichtigt. Die einzelnen Indikatoren werden dichotom kodiert. Für die fünf Facetten der Kategorie „*Wissenschaftlichkeit der Skizze*“ und die aufgabentypspezifischen zusätzlichen Kategorien werden Summen-Scores der einzelnen Facetten gebildet. Sogenannte Schwellenwerte (siehe Kapitel 6.5.2) markieren die Grenze, wann ein Schüler oder eine Schülerin das nächsthöhere Niveau erreicht. Diese Schwellen-

werte stützen sich auf die holistische Beschreibung der Niveaus des Progressionsmodells und hängen von der leicht unterschiedlichen Anzahl Indikatoren pro Facette je Kontext ab.

Ein Beispiel aus dem Aufgabentyp „Beobachtung“ im Kontext „Brausetablette“ mit der Facette „Idealisierung“ ist der folgende Indikator: Wurde weggelassen oder, falls gezeichnet, idealisiert?

- i. die Brausetablette-Dose, Thermosflasche und Wasserflasche
- ii. das Schälchen mit Brausetablette (weggelassen!)
- iii. wurde alles andere weggelassen (Hände etc.) (siehe Anhang Kodiermanual)

6.1.3 Progression und Einbettung der Qualitätsmerkmale ins Kompetenzstrukturmodell

In Kompetenzstrukturmodellen kann die Unterteilung von Kompetenzausprägungen quantitativ festgestellt und können Kompetenzunterschiede zwischen Personen beschrieben werden (Fleischer et al., 2013). In den verschiedenen Niveaus im Progressionsmodell wird beschrieben, über welche Kompetenzausprägung eine Schülerin oder ein Schüler mindestens verfügen muss, um dieses Niveau zu erreichen. Für die theoriebasierten Modellierungen eines Kompetenzstrukturmodells des *wissenschaftlichen Skizzierens* wurden die in der Literatur aufgeführten Kriterien für die Konstruktion von Skizzen berücksichtigt (siehe Kapitel 3.5.2). Das Progressionsmodell (siehe Kapitel 6.2) wurde a priori in einem literaturbasierten Ansatz entwickelt, wobei die Kompetenzen, die ein Lernender für ein bestimmtes Niveau erfüllen muss, aufgeführt wurden. Als Ergänzung zum deduktiven, forschungsbasierten Ansatz (siehe Kapitel 3.5.2) wurden sowohl eine Expertenbefragung (siehe Kapitel 7.2) durchgeführt als auch Erkenntnisse aus bereits vorliegenden Schülerskizzen (Cavelti et al., 2018) berücksichtigt.

Die Kompetenzen im Allgemeinen wie auch die Kompetenzausprägungen der einzelnen Niveaus sind holistisch beschrieben (siehe Kapitel 6.2) und werden durch das in Kategorien und Facetten (siehe Tab. 6 und 7) aufgeteilte und nach Kontexten (siehe Tab. 19) aufgetrennte Kodiermanual (siehe Anhang) ergänzt. Mit Hilfe der sogenannten Indikatoren (Beispiel siehe Anhang) werden im Kodiermanual die einzelnen Facetten und Kategorien beschrieben. Die holistische Beschreibung der Niveaus des

Progressionsmodells (siehe Kapitel 6.2) wurde theoriebasiert (siehe Kapitel 3.5.2) entwickelt.

Tab. 8 Literaturbezug der Inhalte der Facetten der Kategorie „Wissenschaftlichkeit der Skizze“

Facette	Literaturbezug der Inhalte der Facetten
Eindeutigkeit	Eindeutigkeit (hat eine einzige Interpretation) (diSessa, 2002) Konventionen (wurden keine Konventionen gebrochen) (diSessa, 2004) Autonomie (aus eigener Kraft nachvollziehbar) (diSessa, 2002) Gewünschte Attribute (diSessa, 2004) Klarheit (Aussagekraft) (Gebre & Polman, 2016) Organisation und Design (Gebre & Polman, 2016) Lesbarkeit (Gebre & Polman, 2016)
Idealisierung	Sparsamkeit (wurden keine unnötigen Elemente genutzt) (diSessa, 2004) Systematik (wurden einfache Regeln befolgt) (diSessa, 2004) Einfachheit (vermeidet unnötige Komplexität) (diSessa, 2002) Metrische Beziehungen (diSessa, 2002) Erkenntnisbezogene Genauigkeit (Vollständigkeit und Genauigkeit) (diSessa, 2002) Realismus (diSessa, 2004)
Vollständigkeit	Vollständigkeit (sind alle relevanten Informationen dargestellt) (diSessa, 2004) Systematik (wurden einfache Regeln befolgt) (diSessa, 2004) Anordnung (die Beziehungen zwischen den Teilen der Repräsentation sind klar dargestellt) (diSessa, 2004) Vollständigkeit (abhängig von Daten, Inhalt, Umfang und Tiefe) (Gebre & Polman, 2016) Übereinstimmung zwischen den einzelnen Teilen (Gebre & Polman, 2016) Verwendung von Pfeilen, um Abfolgen darzustellen (Cavelti et al., 2018)
Abstraktion	Abstraktheit versus Realismus (diSessa, 2004) Relevanz (Gebre & Polman, 2016)
Korrektheit	Präzision (sind die quantitativen Informationen präzise lesbar) (diSessa, 2004) Erkenntnisbezogene Genauigkeit (Vollständigkeit und Genauigkeit) (diSessa, 2002) Verwendung von Pfeilen, um Bewegungen darzustellen (Jee et al., 2014) Unterscheidung von Beobachtung und Erklärung (Klos et al., 2008)

In der Literatur werden verschiedene Qualitätsmerkmale des Skizzierens diskutiert (siehe Kapitel 3.5.2), die auch für die Kompetenz des *wissenschaftlich Skizzierens* Gültigkeit haben. Sie bilden die Grundanforderungen an *wissenschaftliche Skizzen* und somit für die Entwicklung eines Kompetenzstrukturmodells. Die Qualitätsmerkmale werden im Folgenden aufgeführt:

Tab. 8 zeigt eine Zusammenstellung der in der Literatur diskutierten Qualitätsmerkmale des (wissenschaftlicher) Skizzierens und der Darstellung von Repräsentationen (siehe Kapitel 3.6.2), die den einzelnen Facetten der Kategorie der „*Wissenschaftlichkeit der Skizze*“ zugeordnet sind; es ist eine Setzung aufgrund der Überlegungen zur Modellierung zur Kompetenz des *wissenschaftlichen Skizzierens*.

Die Zuordnungen erfolgten entsprechend den erarbeiteten Grundanforderungen an das *wissenschaftliche Skizzieren* (siehe Tab. 4). Tab. 8 ist das Produkt der kognitiven Synthese der literaturbasierten Qualitätsmerkmale und Kriterien des Skizzierens und der Darstellung von Repräsentationen. Literaturbezug und Kompetenzanforderung werden exemplarisch an der Facette „Eindeutigkeit“ aufgezeigt: Die Kriterien Eindeutigkeit (hat eine einzige Interpretation) (diSessa, 2002), Konventionen (wurden keine Konventionen gebrochen) (diSessa, 2004), Autonomie (aus eigener Kraft nachvollziehbar) (diSessa, 2002), Gewünschte Attribute (diSessa, 2004), Klarheit (Aussagekraft) (Gebre & Polman, 2016), Organisation und Design (Gebre & Polman, 2016) und Lesbarkeit (Gebre & Polman, 2016) fließen in die Facette „Eindeutigkeit“ ein, neben „erkennbar, nachvollziehbar, dass das Phänomen/die Phänomene erzeugt bzw. welche Variablen kontrolliert werden“, und bilden dadurch Grundlage für die Kompetenzanforderungen im Progressionsmodell des *wissenschaftlichen Skizzierens* (siehe Kapitel 6.2).

6.2 Progressionsmodelle des wissenschaftlichen Skizzierens

Bei den Kompetenzausprägungen kann, ausgehend vom Kompetenzstrukturmodell, entweder die Problemkomplexität variiert oder die Lösungsqualität von Aufgaben kodiert werden (siehe Kapitel 3.5). Die Kompetenzen des *wissenschaftlichen Skizzierens* wurden für die Kategorien (siehe Tab. 7) „*Wissenschaftlichkeit der Skizze*“ wie auch für die beiden aufgabentypspezifischen Kategorien „Unterschied erfasst und Vergleich dargestellt“ und „Variablen erfasst und vollständig kontrolliert“ einzeln in einer Progression aufgeführt, wobei darauf geachtet wurde, dass höhere Fähigkeiten einer höheren Progressionsstufe entsprechen. Zwei Progressionsstufen beziehungsweise zwei Ausprägungen einer Kompetenz korrespondieren im Idealfall mit einer zunehmenden Kompetenz.

Tab. 9 zeigt das Progressionsmodell für den Aufgabentyp „Beobachtung“. Tab. 10 zeigt das Progressionsmodell für den Aufgabentyp „Vergleich“. Tab. 11 zeigt das Progressionsmodell für den Aufgabentyp „Experimentieren“.

Tab. 9 Progressionsmodell des Aufgabentyps „Beobachtung“

Niveau	Kategorien für „Beobachtung“
	- Wissenschaftlichkeit der Skizze - Eindeutigkeit (eindeutig interpretierbar): relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze eindeutig interpretierbar - Idealisierung: Relevanz/Fokus, keine für die Forschungsfrage irrelevanten Objektmerkmale dargestellt - Vollständigkeit: Vorgänge/Abläufe dargestellt (Bildergeschichte/Nummern/ Verwendung von Pfeilen, um Vorgänge zu zeigen) - Abstraktion(sgrad): Abstraktion der Darstellung, keine Fotografie - fachliche Korrektheit: relevante Proportionen korrekt, relevante Objektmerkmale der Reaktion korrekt
3	- relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze ohne zusätzliche Erklärung eindeutig (eindeutig interpretierbar) - Idealisierung: Relevanz/Fokus, keine irrelevanten Dinge dargestellt, nur Relevantes, fokussiert, falls Text vorhanden, gibt dieser zusätzliche (nicht zeichenbare) Informationen - Vorgänge/Abläufe vollständig dargestellt (Bildergeschichte/ Nummern, Verwendung von Pfeilen, um Vorgänge zu zeigen) - hoher Abstraktionsgrad - größtenteils fachliche Korrektheit, relevante Proportionen größtenteils korrekt dargestellt, relevante Objektmerkmale der Reaktion größtenteils korrekt dargestellt
2	- mindestens relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze größtenteils ohne zusätzliche Erklärung (eindeutig Interpretierbar) - nichts bis höchstens wenig Irrelevantes, gar nicht bis höchstens wenig unfokussiert, falls Text vorhanden, gibt dieser zusätzliche (nicht zeichenbare) Informationen - Vorgänge/Abläufe mindestens größtenteils dargestellt (Bildergeschichte/ Nummern/ Verwendung von Pfeilen, um Vorgänge zu zeigen) - mindestens überwiegend abstrahiert - mindestens teilweise fachlich korrekt, relevante Objektmerkmale der Reaktion mindestens teilweise korrekt dargestellt (nur kleine Fehler oder etwas Fehlendes)
1	- mindestens größtenteils relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, größtenteils Phänomen dargestellt, Skizze mindestens teilweise ohne zusätzliche Erklärung eindeutig (größtenteils eindeutig interpretierbar) - mindestens teilweise Irrelevantes, teilweise unfokussiert, falls Text vorhanden sagen Text und Skizze dasselbe (redundant)

	- Vorgang des Reaktionsverlaufs mindestens teilweise gezeichnet (Pfeile ansatzweise verwendet) teilweise Bildabfolgen - mindestens teilweise abstrahiert (eher Fotografie-ähnlich) - mindestens größtenteils fachlich nicht korrekt (fehlerhaft, Fehlendes)
0	- nicht oder ansatzweise relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze nicht oder ansatzweise eindeutig ohne zusätzliche Erklärung (ansatzweise eindeutig interpretierbar) - nicht oder mindestens teilweise Irrelevantes, unfokussiert, falls Text vorhanden, kann dieser in eine Skizze überführt werden - Reaktionsverlauf nicht oder nur teilweise dargestellt - nicht oder mindestens ansatzweise abstrahiert (Fotografie-ähnlich) - nicht oder nur mindestens ansatzweise korrekt (viele fachliche Fehler und Fehlendes)

Tab. 10 Progressionsmodell des Aufgabentyps „Vergleich“

Niveau	Kategorien für „Vergleich“	
	- Wissenschaftlichkeit der Skizze - Eindeutigkeit (eindeutig interpretierbar): relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze eindeutig interpretierbar - Idealisierung: Relevanz/Fokus, keine für die Forschungsfrage irrelevanten Objektmerkmale dargestellt - Vollständigkeit: Vorgänge/Abläufe dargestellt (Bildergeschichte/Nummern/ Verwendung von Pfeilen, um Vorgänge zu zeigen) - Abstraktion(sgrad): Abstraktion der Darstellung, keine Fotografie - fachliche Korrektheit: relevante Proportionen korrekt, relevante Objektmerkmale der Reaktion korrekt	- Unterschied erfasst und - Vergleich dargestellt
3	- relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze größtenteils ohne zusätzliche Erklärung eindeutig (eindeutig interpretierbar) - Idealisierung: Relevanz/Fokus, keine irrelevanten Dinge dargestellt, nur Relevantes, fokussiert, falls Text vorhanden, gibt dieser zusätzliche (nicht zeichenbare) Informationen - Vorgänge/Abläufe vollständig dargestellt (Bildergeschichte/Nummern, Verwendung von Pfeilen, um Vorgänge zu zeigen) - hoher Abstraktionsgrad - grösstenteils fachliche Korrektheit, relevante Proportionen größtenteils korrekt dargestellt, relevante Objektmerkmale der Reaktion größtenteils korrekt dargestellt	- Vergleich vollständig und korrekt dargestellt (Vergleichsbedingung dargestellt) - Relevante Unterschiede/merkmale vollständig erfasst

2	- mindestens relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze ohne zusätzlich Erklärung (eindeutig Interpretierbar) - nichts bis höchstens wenig Irrelevantes, gar nicht bis höchstens wenig unfokussiert, falls Text vorhanden, gibt dieser zusätzliche (nicht zeichenbare) Informationen - Vorgänge/Abläufe mindestens größtenteils dargestellt (Bildergeschichte/ Nummern/ Verwendung von Pfeilen, um Vorgänge zu zeigen) - mindestens überwiegend abstrahiert - mindestens teilweise fachlich korrekt, relevante Objektmerkmale der Reaktion mindestens teilweise korrekt dargestellt (nur kleine Fehler oder etwas Fehlendes)	- Vergleich vollständig korrekt dargestellt - Relevante Unterschiedsmerkmale mindestens teilweise erfasst
1	- mindestens größtenteils relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, größtenteils Phänomen dargestellt, Skizze mindestens teilweise ohne zusätzliche Erklärung eindeutig (größtenteils eindeutig interpretierbar) - mindestens teilweise Irrelevantes, teilweise unfokussiert, falls Text vorhanden, sagen Text und Skizze dasselbe (redundant) - Vorgang des Reaktionsverlaufs mindestens teilweise gezeichnet (Pfeile ansatzweise verwendet) teilweise Bildabfolgen - mindestens teilweise abstrahiert (eher Fotografie-ähnlich) - mindestens größtenteils fachlich nicht korrekt (fehlerhaft, Fehlendes)	- Vergleichsvariablen erkannt Vergleich unvollständig und mindestens nicht vollkommen korrekt dargestellt - Relevante Unterschiedsmerkmale nicht oder mindestens ansatzweise erfasst
0	- nicht oder ansatzweise relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze nicht oder ansatzweise eindeutig ohne zusätzliche Erklärung (ansatzweise eindeutig interpretierbar) - nicht oder mindestens teilweise Irrelevantes, unfokussiert, falls Text vorhanden, kann dieser in eine Skizze überführt werden - Reaktionsverlauf nicht oder nur teilweise dargestellt - nicht oder mindestens ansatzweise abstrahiert (Fotografie-ähnlich) - nicht oder nur mindestens ansatzweise korrekt (viele fachliche Fehler und Fehlendes)	- Vergleichsvariablen nicht oder mindestens teilweise erkannt, aber nicht vollständig dargestellt - Relevante Unterschiedsmerkmale nicht erfasst

Tab. 11 Progressionsmodell des Aufgabentyps „Experimentieren“

Niveau	Kategorien für „Experimentieren“	
	- Wissenschaftlichkeit der Skizze - Eindeutigkeit (eindeutig interpretierbar): relevante Objektmerkmale zum Erstellen des Phänomens dargestellt (alle Stoffe vorhanden), Skizze eindeutig interpretierbar - Idealisierung: Relevanz/Fokus, keine für die Forschungsfrage irrelevanten Objektmerkmale dargestellt - Vollständigkeit: Vorgehen dargestellt (Ablauf: Verwendung von Pfeilen/Nummern), Phänomen nicht dargestellt keine irrelevanten Objektmerkmale dargestellt - Abstraktion(sgrad): Abstraktion der Darstellung, keine Fotografie - fachliche Korrektheit: relevante Proportionen korrekt dargestellt, relevante Objektmerkmale des Experiments korrekt dargestellt	- Variablen erfasst und vollständig kontrolliert - (Experimentieren)
3	- relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt. Skizze größtenteils eindeutig ohne zusätzliche Erklärung (eindeutig interpretierbar) - Idealisierung: Relevanz/Fokus, keine irrelevanten Dinge dargestellt, nur Relevantes, fokussiert, falls Text vorhanden, gibt dieser zusätzliche (nicht zeichenbare) Informationen - Vorgehen vollständig dargestellt (Ablauf) (was wird zu was gegeben?), Phänomen/Reaktion nicht dargestellt - hoher Abstraktionsgrad - fachliche Korrektheit (relevante Proportionen korrekt dargestellt, relevante Objektmerkmale des Experiments korrekt dargestellt)	- Beide der zu kontrollierenden Variablen erfasst und dargestellt - Beide Variablen vollständig kontrolliert
2	- mindestens relevante Objektmerkmale zum Erstellen des Phänomens dargestellt (alle Stoffe vorhanden), Skizze eindeutig interpretierbar - nicht bis höchstens wenig Irrelevantes, gar nicht bis höchstens wenig unfokussiert, falls Text vorhanden, gibt dieser zusätzliche (nicht zeichenbare) Informationen - Vorgehen mindestens größtenteils dargestellt (Ablauf) (was wird zu was gegeben?) - mindestens überwiegend abstrahiert, Phänomen/Reaktion nicht dargestellt - mindestens teilweise fachliche Korrektheit (relevante Proportionen korrekt dargestellt, relevante Objektmerkmale des Experiments mindestens teilweise korrekt dargestellt, kleinere Fehler oder Fehlendes)	- Beide der zu kontrollierenden Variablen erfasst und dargestellt - Eine der zu kontrollierenden Variablen mindestens vollständig kontrolliert, zweite der zu kontrollierenden Variablen nicht oder mindestens

		ansatzweise kontrolliert
1	- mindestens relevante Objektmerkmale zum Erstellen des Phänomens größtenteils vollständig dargestellt (alle Stoffe vorhanden), Skizze mindestens größtenteils eindeutig interpretierbar. - mindestens teilweise Irrelevantes, mindestens teilweise unfokussiert, falls Text vorhanden, sagen höchstens Text und Skizze dasselbe (redundant). - Vorgehen mindestens teilweise dargestellt (Ablauf) (was wird zu was gegeben?) - mindestens teilweise abstrahiert - mindestens ansatzweise fachliche Korrektheit (relevante Proportionen, relevante Objektmerkmale des Experiments ansatzweise korrekt dargestellt (fehlerhaft oder Fehlendes)), Phänomen mindestens angedeutet dargestellt	- Beide der zu kontrollierenden Variablen erfasst und dargestellt - Beide Variablen nicht oder mindestens ansatzweise kontrolliert
0	- nicht oder mindestens ansatzweise relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Skizze mindestens ansatzweise eindeutig ohne zusätzliche Erklärung (ansatzweise eindeutig interpretierbar) - mindestens teilweise Irrelevantes, unfokussiert, falls Text vorhanden, kann dieser in eine Skizze überführt werden - Vorgehen nicht oder mindestens ansatzweise dargestellt (Ablauf) (was wird zu was gegeben?) - nicht oder mindestens ansatzweise abstrahiert (Fotografie-ähnlich) - nicht oder mindestens ansatzweise korrekt (viele fachliche Fehler und Fehlendes) Objektmerkmale des Experiments nicht oder mindestens ansatzweise korrekt dargestellt, Phänomen mindestens ansatzweise angedeutet dargestellt.	- keine oder eine der zu kontrollierenden Variablen erfasst und dargestellt - Zweite der zu kontrollierenden Variablen nicht erfasst und nicht dargestellt - Variablen nicht kontrolliert

Das hier entwickelte Kompetenzstrukturmodell zum *wissenschaftlichen Skizzieren* enthält vier Niveaus. Hope's Framework (Hope, 2008) (siehe Kapitel 3.5.1) verfügt über fünf Stufen. Die höchste Stufe („interaktive“) umfasst Kompetenzen wie Reflexion und interaktive Vernetzung. Diese Art von Kompetenzen werden hier nicht direkt erhoben und können dadurch auch nicht gemessen werden.

Das Niveau 0 der Kompetenz des *wissenschaftlichen Skizzierens* deckt neben der ersten Stufe von Hope's Framework auch Bereiche seiner zweiten Stufe ab. Niveau 1 der Kompetenz des *wissenschaftlichen Skizzierens* kann mit dem Niveau 2 von Hope's Framework in Verbindung gebracht werden. „single-draw“ zeigt Ähnlichkeit mit den holistischen Beschreibungen im Kompetenzstrukturmodell. Ebenso zeigen die beiden

höchsten Niveau des Kompetenzstrukturmodells des *wissenschaftlichen Skizzierens* ähnliche Ansätze wie die Stufen 3 und 4 von Hope's Framework. Die Studierenden müssen ab dem Niveau 2 des Kompetenzstrukturmodells des *wissenschaftlichen Skizzierens* mehrere Dinge gleichzeitig berücksichtigen können (siehe Abb. 3).

HOPE	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
wissenschaftliches Skizzieren	Niveau 0	Niveau 1	Niveau 2	Niveau 3	

Abb. 3 Vergleich der Niveaus von Hope's Framework den Niveaus des wissenschaftlichen Skizzierens

6.3 Entwicklung Aufgabenset

Grundlage der Itemkonstruktion (Aufgabenset) bildete das Kompetenzmodell (siehe Kapitel 6.2) mit den drei Aufgabentypen. Tab. 12 zeigt eine Gliederung der Aufgabenstellung, der Aufgabentypen und der verwendeten Kontexte. Die Kontexte umfassen das Schmelzverhalten von Eis (nachfolgend „Eis schmelzen“ genannt), chemische Reaktionen (Lösen von Brausetabletten, nachfolgend „Brausetablette“ genannt) und das Visualisieren von Säure-Basen-Reaktionen mit Indikatoren (nachfolgend „Säure und Base“ genannt) sowie die im nur in der Vorpilotierung verwendeten Kontexte Chromatographie und Kohlendioxid.

Der Inhaltsvalidität bei der Testentwicklung wird durch fachliche Überlegungen zum bereits vorhandenen Können der Lernenden und durch subjektive Einschätzungen (Validierung durch Lehrpersonen der Sekundarstufe I (siehe Kapitel 6.2) Rechnung getragen (Bortz & Döring, 2007).

Die Aufgabenstellung der Items ist eine bildliche Darstellung (*wissenschaftliche Skizze*) des Phänomens/ der Phänomene oder der Planung eines Experiments, also des Inhalts der jeweiligen Aufgabe. Die Items rufen dabei kein Vorwissen ab. Kontext und Aufgabentyp bestimmen dabei den Umfang eines Items und damit dessen Schwierigkeitsgrad.

Mit jedem Item wird die Kompetenz des Progressionsmodells gemessen, wodurch eine Aussage über die jeweils erreichte Kompetenzstufe gemacht werden kann. Ein Aufgabenbeispiel zum Aufgabentyp „Vergleich“ im Kontext „Brausetablette“ ist das

Beobachten des unterschiedlichen Auflöseverhaltens von Brausetablette und Smartie in Mineralwasser.

Es wurden curriculare Kontexte (siehe Tab. 12 und 19) gewählt, zu denen Items in allen drei Aufgabentypen entwickelt werden konnten. Der Aufgabenstamm (Vignette) wurde für alle Items identisch formuliert. Er dient dazu, in den Aufgabenkontext einzuleiten und die Schülerinnen und Schüler anzuregen, sich mit den Items auseinanderzusetzen. Um sicherzustellen, dass verschiedene Beobachter unter den gleichen Bedingungen zu demselben Resultat gelangen (Sommer et al., 2018), wurden kurze Videosequenzen, sogenannte Video-Vignetten, eingesetzt. Durch einen Theorie-Input zur Kompetenz des *wissenschaftlichen Skizzierens* (siehe Kapitel 6.3.3) wurden den Lernenden im Vorfeld der Erhebung die Grundregeln zur Erstellung *wissenschaftlicher Skizzen* vermittelt.

Tab. 12 Kontexte und Aufgabenstellungen der Vorpilotierung

Aufgabentyp	Beobachtung	Vergleich	Experimentieren
Aufgabenstellung	Experimentieranordnung und Experimentierablauf skizzieren	erzeugtes Phänomen skizzieren	erzeugte Phänomene vergleichen und skizzieren
Kontext			
Chromatographie	Chromatographie mit schwarzem Farbstift in Wasser	Chromatographie mit schwarzem Farbstift in Ethanol und Wasser	Einfluss der Laufmittel Ethanol und Wasser auf das Laufverhalten verschiedener Farbstifte (Farben) untersuchen
Brausetablette	Lösen einer Brausetablette in Wasser bei Raumtemperatur	Lösen einer Brausetablette und eines Smartie in Mineralwasser bei Raumtemperatur	Einfluss unterschiedlicher Temperaturen und Wassermengen auf das Löseverhalten einer Brausetablette untersuchen
Eis schmelzen	Schmelzen eines mit Natriumchlorid bestreuten Eiswürfels	Schmelzen eines eingefärbten Eiswürfels in Salz und Süßwasser	Einfluss unterschiedlicher Wassermengen und Wasserarten (Salz- bzw. Süßwasser, unterschiedlicher Salzgehalt) auf das Schmelzen eines Eiswürfels untersuchen
Säuren und Basen	Reaktion von Essigsäure mit	Reaktion von Wasser mit zwei	Einfluss unterschiedlicher Schwarzteemengen und

	Universalindikator und Natronlauge	unterschiedlichen Indikatoren (Cresolphthalein und Bromthymolblau) und Natronlauge	unterschiedlicher Zugaben (Zitronensaft und Kernseife) auf die Farbe des Schwarztee untersuchen
Kohlenstoffdioxid	Reaktion von Backpulver und Essig.	Reaktion von Essigsäure mit Universalindikator und Natronlauge bzw. Sodalösung	Einfluss unterschiedlicher Mengen an Backpulver und Zitronensäure sowie unterschiedlicher Wassermengen auf die Gasentwicklung untersuchen

Die Itemstämme (siehe Kapitel 6.3.1, Abb. 4 und Anhang), ergänzt durch die jeweilige Video-Vignette, enthalten alle Informationen, die zur Bearbeitung der Aufgaben (Items) nötig sind. Das offene Itemformat beinhaltet nach dem Arbeitsauftrag einen Platzhalter für die *wissenschaftliche Skizze* mit dem Titel „Hier ist Platz für deine Skizze“. Ergänzend wurde ein Kasten mit dem Titel „Legende“ im Platzhalter gesetzt, um die Schülerinnen und Schüler zur Verwendung einer Legende anzuregen. Itemstamm und vorangegangener Theorieinput zum *wissenschaftlichen Skizzieren*, sprechen klar gegen die Verwendung weiterer sprachlicher Erklärungen oder Ergänzungen im Antwortformat. Die Itemstämme wurden kurz und sprachlich einfach gehalten, um den Einfluss des Leseverständnisses und der Lesegeschwindigkeit so gering wie möglich zu halten. Ein Beispiel eines Aufgabendossiers ist im Anhang abgebildet.

Bei den Antwortmöglichkeiten des dem Aufgabenformat angeschlossenen Multiple-Select Tests (siehe Kapitel 6.3.4) wurde beachtet, dass sowohl die Distraktoren als auch der Attraktor einen gleichmäßigen Satzbau und Textumfang aufweisen. Zur Überprüfung der Testqualität wurden die entwickelten Testinstrumente in einer Vorpilottstudie evaluiert. Die Erkenntnisse und die daraus folgende Weiterentwicklung des Aufgabensets werden im Kapitel 6.4.3 vorgestellt.

Ausgehend von einer Ideensammlung wurden Kontexte fokussiert, die auf alle drei Aufgabentypen anwendbar sind. Es wurde darauf geachtet, dass die Kontexte dem Anfangsunterricht der Chemie und ebenso dem Lehrplan 21 (D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz, 2016b) entsprachen. Außerdem waren für die Auswahl folgende Kriterien relevant:

Für den Aufgabentyp „Beobachtung“ steht ein Phänomen im Zentrum. Dieses soll deutlich erkennbar und innerhalb eines Zeitraums von höchstens drei Minuten be-

obachtbar sein. Der Verlauf des Phänomens lässt sich in einer Skizze abbilden. Bspw. kann das Lösen von Zucker zwar beobachtet werden, ist aber schwer als Skizze abbildbar, außer man arbeitet mit Modellen wie dem Teilchenmodell. Auf Modellarbeit wurde in dieser Arbeit bewusst verzichtet.

Beim Aufgabentyp „Vergleich“ sollen beide Phänomene deutlich erkennbar sein, sich in einer Skizze abbilden lassen und innerhalb eines Zeitraums von höchstens drei Minuten beobachtbar sein. Die beiden Phänomene unterscheiden sich in mindestens einem Merkmal, das deutlich zu erkennen ist.

Beim Aufgabentyp „Experimentieren“ sollen die beiden zu untersuchenden Zusammenhänge, beziehungsweise die Variablen, deutlich voneinander unterscheidbar und klar ersichtlich sein sowie sich in einer Skizze abbilden lassen.

Die fachlichen Kontexte des Aufgabensets basieren auf dem Anfangsunterricht der Chemie im Bereich „Chromatographie“, „Schmelzverhalten von Eis“, „chemische Reaktionen“ (Lösen von Brausetabletten und Kohlenstoffdioxidentstehung), und „Visualisieren von Säure-Base-Reaktionen mit Indikatoren“ und sind stark an die curricularen Inhalte des in der Schweiz geltenden Lehrplans 21 (D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz, 2016b) angelehnt. Dem breiten Schwierigkeitsspektrum der Items (Jenßen et al., 2015) wurde durch die Verwendung unterschiedlich vertrauter Kontexte, die unterschiedliche Anzahl an beobachtbaren Abläufen (beispielsweise Bewegung der Brausetablette) bzw. beobachtbaren Unterschieden zwischen zwei Phänomenen (z.B. unterschiedliche Blasengröße im Aufgabentyp „Vergleich“ des Kontexts „Brausetablette“) Rechnung getragen. Insgesamt wurden fünfzehn Items für den Kompetenzbereich *wissenschaftliches Skizzieren* entwickelt. Nach der Durchführung der Vorpilotstudie wurden diejenigen Items, die die beste Modellpassung aufwiesen, ausgewählt und in der Hauptstudie zur Überprüfung der Forschungsfragen eingesetzt. Tab. 12 zeigt die Phänomene bzw. Untersuchungsaspekte nach Aufgabentyp und Kontext getrennt.

6.3.1 Aufgabenstruktur und Aufgabenstamm (Vignette)

Bei der Konstruktion wurde aufgrund der heterogenen Zielgruppe auf fachlich korrekte, einfach verständliche Sprache mit möglichst geringem Textumfang Wert gelegt (Rost, 2004). Dies dient zur Vermeidung von sprachlichen Konstrukt-irrelevanten

Itemschwierigkeiten auf die Kontexte. Die Texte wurden bei allen Items standardisiert.

Der Beobachtungsfokus (siehe Kapitel 3.2.2.1) in der Aufgabenstellung wurde durch kategoriengeleitete Aufträge auf das Relevante gerichtet (Hild et al., 2015). Nur wenn man weiß, wo man hinschauen soll, sieht man das, was man sehen soll. Der Beobachtungsfokus wurde in der Aufgabenstellung bewusst wiederholt.

Um Trigger zu vermeiden, wurde auf gezeichnete Bilder des Versuchsaufbaus oder der Materialien in der Aufgabenstellung verzichtet. Ein Foto mit Beschriftung wurde zur Sicherstellung der Begriffsbezeichnung der verwendenden Materialien und Chemikalien in den Aufgabenstamm aufgenommen.

Eine sogenannte „Vignette“ wurde bereits beim Projekt ExKoNawi (Bonetti et al., 2017; Gut et al., 2014) wie auch bei Hild et al. (2019) angewendet. Sie ermöglicht den Schülerinnen und Schülern einen einfacheren Zugang und ein besseres Verständnis der Aufgabenstellung. Die Rahmengeschichte beschreibt die Situation (siehe Abb. 4) und hilft den Lernenden, sich mit der Aufgabenstellung zu identifizieren.

Sophie und Tim sprechen französisch und leben in Lausanne. Sie verstehen kaum Deutsch. Die beiden haben in einem Experiment herausgefunden, was mit einer Brausetablette geschieht, wenn diese mit Wasser in Berührung kommt. Sie haben keine Skizzen angefertigt. Zeichne für Sophie und Tim. Beachte dabei: Deine Skizze soll von Sophie und Tim verstanden werden, auch wenn sie kein Deutsch verstehen.
--

Abb. 4 Beispiel einer Vignette

6.3.2 Video-Vignetten

Um sicherzustellen, dass verschiedene Beobachter unter den gleichen Bedingungen zu demselben Resultat gelangen (Pfeifer, 2002), wurden den Lernenden Video-Vignetten vorgelegt. Dadurch wurden Fehler oder Unterschiede beim Experimentiervorgang ausgeschlossen. Die Video-Vignetten wurden mit der Kamera eines iPad® und iPhone® gefilmt und mit dem Programm iMovie® geschnitten. Auf eine Tonspur wurde verzichtet. Die Filme mit einer Länge von rund einer Minute wurden mit einem Titel sprachlich ergänzt. Dieser wird während der ersten drei Sekunden eingeblendet.

6.3.3 Theorieblatt „Regeln des wissenschaftlichen Skizzierens“

Zwecks Schulung der Lernenden vor der Erhebung (siehe Kapitel 6.4) wurde ein Theorieblatt „Regeln des *wissenschaftlichen Skizzierens*“ (siehe Anhang) entwickelt, das sich auf das erarbeitete Progressionsmodell (siehe Kapitel 6.2) stützt und in drei Bereiche unterteilt ist: einen allgemeingültigen Bereich, einen Bereich, der für die beiden Aufgabentypen „Beobachtung“ und „Vergleich“ gilt, und einen Bereich, der für den Aufgabentyp „Experimentieren“ gültig ist. Allgemein gültige Regeln sind keine Erklärungen aufzuschreiben, eine Legende zu verwenden und ein Hinweis, dass es irrelevant ist, ob die Skizze in den Augen der Lernenden als schön gilt.

Auch das Theorieblatt (siehe Anhang) wurde in einer einfach verständlichen Sprache mit geringem Textumfang gehalten (Rost, 2004). Dadurch wurde der Heterogenität der Zielgruppe Rechnung getragen. Das Theorieblatt wurde in Anlehnung an die Regeln im Dokumentier-Führerschein (Cavelti et al., 2018) entwickelt.

6.3.4 Entwicklung Fach- und Methodenwissenstest

Der Multiple-Select Test ist ein Leistungstest (Rost, 2004) und dient dazu, das fachliche und methodische Verständnis der Lernenden zu erfassen. Zu jedem Item wurde ein Set an Multiple-Select Aufgaben (Beispiel siehe Anhang) entwickelt, welche die Auswahl mehrerer Antworten zulassen und dadurch die Ratewahrscheinlichkeit minimieren. Die Aufgaben zielen auf die Erhebung von Fach- und Methodenwissen zu den aufgaben- und kontextspezifischen fachlichen Inhalten ab, die bei der Ermittlung der Kompetenz des *wissenschaftlichen Skizzierens* relevant sind. Bei den Aufgabentypen „Beobachtung“ und „Vergleich“ steht neben der Methodenkompetenz des Beobachtens und Vergleichens vor allem die Fachkompetenz im Zentrum. Es handelt sich einerseits um kategorische Fragen wie „ist passiert/ ist vorhanden/ ist nicht vorhanden“, andererseits beim Aufgabentyp „Vergleich“ zusätzlich um graduelle Fragen wie „ist schneller bei.../ grösser als bei...“. Beim Aufgabentyp „Experimentieren“ zielen die Fragen auf die Variablenkontrolle ab, wie „um den Einfluss von Variable 1 zu untersuchen, vergleicht man folgende Becher-Paare...“. Diese Aufgaben erheben die Methodenkompetenz des Experimentierens im Sinne der Untersuchung von Zusammenhängen. Die Fragenstruktur der Aufgabentypen wurde über alle Kontexte beibehalten.

Die Multiple-Select Aufgaben für den Aufgabentyp „Beobachtung“ umfassen drei Fragen (siehe Tab. 13). Das Antwortformat enthält immer drei Antworten, wechselnd zwischen zwei korrekten Antworten und zwei Distraktoren.

Tab. 13 Fragenstruktur des Aufgabentyps „Beobachtung“

Frage 1	Der Fokus wird nochmals aufgegriffen. Das Antwortformat beinhaltet eine korrekte Antwort und zwei Distraktoren, wobei einer davon die Antwort „geschieht nichts“ enthält.
Frage 2	Es wird nach dem Objekt gefragt: Was geschieht damit oder was passiert genau damit? (Fokus präzises Beobachten).
Frage 3	Es wird nochmals nach dem Objekt in der Flüssigkeit gefragt. Was passiert genau damit? (Fokus präzises Beobachten).

Die Multiple-Select Aufgaben für den Aufgabentyp „Vergleich“ umfassen drei Fragen (siehe Tab. 14). Das Antwortformat enthält immer drei Antworten, wechselnd zwischen zwei korrekten Antworten und zwei Distraktoren.

Tab. 14 Fragenstruktur des Aufgabentyps „Vergleich“

Frage 1	Der Fokus des ersten „Phänomens“ wird aufgegriffen. Das Antwortformat beinhaltet eine korrekte Antwort und zwei Distraktoren, wobei einer davon die Antwort „geschieht nichts“ enthält.
Frage 2	Der Fokus des zweiten „Phänomens“ wird aufgegriffen. Das Antwortformat beinhaltet eine korrekte Antwort und zwei Distraktoren, wobei einer davon die Antwort „geschieht nichts“ enthält.
Frage 3	Eine kategorische Frage wie: „Wo passiert es?“, „Was passiert genau?“ wird bei verschiedenen Objekten oder unterschiedlichen Reaktionen gestellt. Bei gleicher Reaktion werden graduelle Fragen wie „Wo ist die Reaktion schneller?“ nach Reihenfolgsunterschieden gestellt.
Frage 4	analog Frage 3

Die Multiple-Select Aufgaben für den Aufgabentyp „Experimentieren“ umfassen zwei Fragen mit je sechs Antwortmöglichkeiten (alle Kombinationen von zwei aus vier Bechern), wobei es sich um zwei korrekte Antworten und vier Distraktoren handelt (siehe Tab. 15).

Tab. 15 Fragenstruktur des Aufgabentyps „Experimentieren“

Frage 1	Um den Einfluss von Variable 1 zu untersuchen, vergleicht man folgende Becher-Paare...
Frage 2	Um den Einfluss von Variable 2 zu untersuchen, vergleicht man folgende Becher-Paare...

6.3.5 Instrumente der externen Validierung

Neben der Kompetenzerfassung des *wissenschaftlichen Skizzierens* und den Multiple-Select Aufgaben wurden in dieser Studie drei Tests zur Erfassung zusätzlicher Kontrollvariablen eingesetzt. Dabei handelte es sich um den kognitiven Fähigkeitstest KFT (Heller & Perleth, 2000), von welchem die nonverbale Skala (N3) verwendet wurde, den Lesegeschwindigkeits- und Leseverstehenstest (LGVT) von Schneider, Schlagmüller und Ennemoser (2007) und einen Test zur Erfassung des Strategiewissens zum Experimentieren (NAW) (Koenen, 2014; Mannel, 2011). Zudem wurde die Anzahl Sprachen, welche die Schülerinnen und Schüler zu Hause sprechen (Deutsch, eine oder zwei Fremdsprachen) und das Geschlecht erhoben.

Der Test zur naturwissenschaftlichen Arbeitsweise (dem Strategiewissen beim Experimentieren, NAW) von Mannel (2011) und Koenen umfasst vierzig Fragen aus den Kategorien „Idee“, „Beobachtung“ und „Ergebnis“. Für die Testung wurden jeweils drei Fragen jeder Kategorie ausgewählt. Auswahlkriterium war der Schwierigkeitsgrad der Fragen, wobei je eine leichte, mittelschwere und schwierige Frage gewählt wurden. Der Schwierigkeitsgrad war zuvor in der Pilot- und Hauptstudie des Projekts ExKoNawi mittels Rasch-Analyse ermittelt worden (Bonetti et al., 2017, 2018).

Zur Erfassung der kognitiven Fähigkeiten wurde der kognitive Fähigkeitstest (KFT) von Heller und Perleth (2000) eingesetzt. Dieser Test ist ein Diagnoseinstrument zur Beurteilung schulleistungsrelevanter Dimensionen kognitiver Fähigkeiten von Schülerinnen und Schülern der 4.-13. Jahrgangsstufe aller Schulstufen (Primar bis Ende 6. Jahrgangsstufe, Sekundarstufe I 6.-9. Jahrgangsstufe, Gymnasien 8.-12. Jahrgangsstufe). In Deutschland erfolgt die Trennung der Schulstufen bereits ab dem 4. Schuljahr in Hauptschulen, Realschulen und Gymnasien. In der Arbeit wurde die Subskala zum nonverbal-figuralen Denken (N3) eingesetzt, um die Fähigkeiten des nonverbalen Denkens zu erfassen.

Die Lesekompetenz der Schülerinnen und Schüler ist für die Bearbeitung der Textaufgaben, aber auch der Multiple-Select Fragen eine wichtige Voraussetzung, da die Aufgabenstellung verstanden werden muss, um die Aufgaben korrekt zu lösen bzw. beantworten die Fragen richtig zu beantworten. In der Annahme, dass sich die Lesekompetenz aus Lesegeschwindigkeit und Leseverständnis zusammensetzt, wurden in dieser Studie beide Komponenten mit Hilfe des Lesegeschwindigkeits- und -verständnistests (LGVT 6-12) von Schneider, Schlagmüller & Ennemoser (2007) in den Jahrgangsstufen 7 und 9 ermittelt. Hier liegen für die Klassenstufen 6 - 9 aller Schularten (außer Sonderschule) gesicherte Prozentrangnormen für die Bewertung (Anzahl korrekter Unterstreichungen sowie Anzahl gelesener Wörter) vor.

6.4 Erhebung

6.4.1 Stichprobe und Testdesign

Um die Anonymität der Schülerinnen und Schüler zu gewährleisten, wurden die eingesetzten Testhefte mit einem Schülercode versehen. Die Schülerinnen und Schüler erhielten Etiketten zum Aufkleben auf die einzelnen Testhefte, versehen mit diesem Schülercode, der sich wie folgt zusammensetzt:

Studie			Jahrgangsstufe		Schülernummer
V	P	H	1-12 (7. Jahrgangsstufe)	13-24 (9. Jahrgangsstufe)	fortlaufend nummeriert z. B. 04

Bspw.: H_01_02

Abb. 5 Konstruktion des individuellen Schülercodes

Anhand des Codes können Rückschlüsse auf Schulform, Schule und Jahrgangsstufe gezogen werden. Ein Rückschluss auf einen individuellen Schüler und eine individuelle Schülerin ist dagegen nicht möglich.

6.4.2 Vorpilotstudie

Zweck der Vorpilotierung war die Erprobung des Aufgabensets, der unterschiedlichen Kontexte sowie der Verständlichkeit der Aufgabentexte. Das Aufgabenset bestand aus der Video-Vignette, dem Auftrag zur Erstellung der *wissenschaftlichen Skizzen* mittels Vignette und einem Set an Multiple-Select-Fragen.

An der Vorpilotstudie (März und April 2019) nahmen 77 Schülerinnen und Schüler aus den Kantonen Zürich und Thurgau teil. In der Schweiz gibt es insgesamt vier

unterschiedliche Anforderungsniveaus in den Jahrgangsstufen 7 - 9: Gymnasium, Sek A, Sek B und Sek C. Nach dem Gymnasium ist das Niveau Sek A das nächsthöhere Anforderungsniveau und entspricht ungefähr dem Niveau der Realschule in Deutschland. Die Schülerinnen und Schüler im Niveau Sek B verfügen über geringere kognitive Fähigkeiten. Dieses Niveau entspricht ungefähr dem Niveau der Hauptschule im deutschen Bildungssystem. Die Lernenden, die in Niveau C unterrichtet werden, benötigen häufig zusätzliche, individuelle Förderung. Das Niveau Sek C liegt daher unter dem Niveau Sek B. In der Schweiz werden auf der Sekundarstufe I je nach Schule die Niveaus getrennt oder in einer niveaudurchmischten Klasse mit dem Niveau Sek AB unterrichtet. Häufig werden in solche Klassen auch Lernende integriert, deren Niveau demjenigen der Sek C entspricht. Auf reine Niveau Sek C Klassen wurde in der Studie ganz verzichtet, da diese Klassen, selten sind und oftmals weniger als zehn Lernenden aufweisen. Die Stichprobe bestand aus Schülerinnen und Schülern der 7. und 9. Jahrgangsstufe der Niveaus Sek A, Sek B und Sek AB sowie der 9. Jahrgangsstufe Gymnasium.

Tab. 16 Units in der Vorpilotstudie

Aufgabentyp	Beobachtung	Vergleich	Experimentieren
Unit 1	Chromatographie	Brausetablette	Schmelzen von Eis
Unit 2	Kohlenstoffdioxid	Chromatographie	Brausetablette
Unit 3	Säure und Basen	Kohlenstoffdioxid	Chromatographie
Unit 4	Schmelzen von Eis	Säure und Basen	Kohlenstoffdioxid
Unit 5	Brausetablette	Schmelzen von Eis	Säure und Basen

Der eingesetzte Test zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* wurde in einem Multi-Matrix-Design angelegt, sodass nicht jeder Schüler und jede Schülerin alle Items aus dem gesamten Kompetenztest bearbeitete. Für die Datenerhebung der Vorpilotierung wurden fünf Units zusammengestellt. Jede Unit beinhaltete ein Item jedes Aufgabentyps, aus drei unterschiedlichen Kontexten stammend (siehe Tab. 16). Jeder Lernende bearbeitete zwei Units, wobei die Schülerinnen und Schüler einer Klasse die gleichen Units erhielten (siehe Tab. 17).

Tab. 17 Rotationsschema Vorpilottstudie

Klasse 1	Units 1, 2
Klasse 2	Units 3, 4
Klasse 3	Units 3, 5
Klasse 4	Units 1, 4

Nach einer Einführung in die Regeln des wissenschaftlichen Skizzierens mittels Theorieblatt (siehe Kapitel 6.3.3) wurden die Units bearbeitet, wobei bei jedem Item das Vorgehen gleich war (siehe Tab. 18). Erst lasen die Lernenden den Einführungstext (Vignette) und schauten sich die Video-Vignette (siehe Kapitel 6.3.2) ein erstes Mal im Plenum an. Die eine Hälfte der Klasse erstellte anschließend die *wissenschaftliche Skizze*, während die andere die Multiple-Select Fragen (siehe Kapitel 6.3.4) bearbeitete, wodurch die Unterschiede in der Reihenfolge der Bearbeitung erfasst wurden. Nach drei Minuten wurde die Video-Vignette erneut gezeigt und im Anschluss daran wurden die Multiple-Select Fragen beantwortet und die *wissenschaftliche Skizze* erstellt. Pro Item standen zehn Minuten zur Verfügung. Zur Überprüfung des Textverständnisses unterstrichen Lernende diejenigen Wörter, die sie nicht verstanden. Die Testung fand an einem Tag statt und umfasste 65-70 Minuten (siehe Tab. 18).

Tab. 18 Erhebungsmanual Vorpilottierung für ein Item

Zeit	Gruppe 1	Gruppe 2
2'	Lektüre Einführungstext	Lektüre Einführungstext
1'	Video-Vignette	Video-Vignette
3'	Skizze	Multiple-Select Fragen
1'	Video-Vignette	Video-Vignette
3'	Multiple-Select Fragen	Skizze

Unterschiede durch die Reihenfolge der Bearbeitung (zuerst Erstellung der Skizze oder zuerst Bearbeitung der Multiple-Select Fragen) konnten nach der Evaluation der Resultate der Vorpilottierung ausgeschlossen werden. Durch das Befolgen des Erhebungsmanuals konnte die Durchführungsobjektivität gewährleistet werden.

6.4.3 Weiterentwicklung Aufgabenset nach Evaluation Vorpilotierung

Nach der Evaluation der Vorpilotierung wurde das Aufgabenset überarbeitet. In den Skizzen der Lernenden, vor allem im Aufgabentyp „Vergleich“, konnte oftmals nicht nachvollzogen werden, um welchen Stoff es sich handelt. Es fehlten außerdem Angaben dazu, was bspw. zur Erzeugung des Phänomens hinzugefügt wurde. Durch die Verwendung einer Legende oder Beschriftungen direkt in der Skizze hätte die Nachvollziehbarkeit der Skizzen erhöht werden können. Das Theorieblatt „Regeln des wissenschaftlichen Skizzierens“ (siehe Kapitel 6.3.3) wurde somit um den Punkt „Verwende eine Legende“ ergänzt. Auch das Kompetenzstrukturmodell (siehe Kapitel 3.5) wurde um diesen Punkt „Verwendung einer Legende“ und um einen weiteren Punkt „Verwendung von Pfeilen, um Abläufe darzustellen“ ergänzt.

Die Testhefte wurden optimiert und sprachlich angepasst. Sprachliche Anpassungen wurden bei der Formulierung des Auftrags beim Aufgabentyp „Experimentieren“ vorgenommen, wodurch der Fokus der Lernenden bewusster auf die Planung des Experiments gelenkt wurde. Weitere sprachliche Anpassungen dienten der Präzision der Multiple-Select Fragen. Bspw. wurde das Wort „Schlieren“ genauer erklärt und auch der Ort der Entstehung der Blasen mit dem Begriff „vor allem“ ergänzt.

Tab. 19 Schwierigkeiten bei den Experimenten und Reaktionen oder den Skizzen

Aufgabentyp	Beobachtung	Vergleich	Experimentieren
Kontext			
Chromatographie	- teilweise ergänzender Text - selten mehrere Bilder des Phänomens - viele Skizzen sehr ähnlich - kaum Varianz	- Vergleich dargestellt - Unterschied erfasst - selten mehrere Bilder der Reaktion - viele Skizzen sehr ähnlich - kaum Varianz - Legende oftmals vorhanden	- Variable Stiftfarbe sehr häufig erkannt - Variable Flüssigkeit sehr häufig erkannt - Darstellung sehr unterschiedlich - Legende oftmals vorhanden - kaum Varianz
Brause-tablette	- mehrere Bilder des Phänomens, jedoch nicht alle Details dargestellt. - alle Niveaus belegt,	- Vergleich „Sprudeln“ dargestellt - Vergleich „Lösen“ oftmals nicht dargestellt	- Variablen Wassermenge oftmals erkannt

	jedoch Ballung im Niveau 2	- teilweise mehrere Bilder der Reaktion - teilweise mit Text ergänzt	- Variable Wassertemperatur häufig erkannt
Eis schmelzen	- Phänomen dargestellt - detaillierter Schmelzvorgang selten gezeichnet - große Unterschiede in den Skizzen	- Vergleich „Schmelzen“ dargestellt - Vergleich „Farbänderung“ teilweise dargestellt - große Unterschiede in den Skizzen - Kodierung einfach	- Variable Wassermenge teilweise erkannt - Variable Salz-/Süßwasser häufig erkannt - Kodierung gut machbar
Säuren und Basen	- Details „Schlieren“ teilweise gezeichnet - teilweise abstrakt, teilweise sehr exakt - Farbänderung teilweise nicht nachvollziehbar - große Unterschiede in den Skizzen - ohne Legende/Test kaum nachvollziehbar - Kodierung einfach	- teilweise sehr exakt, teilweise unklar dargestellt - Vergleich „Farbunterschiede“ oftmals erkannt - Vergleich „Phänomen“ teilweise erkannt - Details selten dargestellt - große Unterschiede in den Skizzen	- Variable Teemenge erkannt - Variable Zugabe häufig erkannt - teilweise abstrahiert - große Unterschiede in den Skizzen
Kohlenstoffdioxid	- Phänomen oftmals dargestellt - selten mehrere Bilder der Reaktion - kaum Varianz - teilweise sehr fotoähnlich - häufig mit Text ergänzt	- Vergleich „Regenbogen“ und „Gasentwicklung“ oftmals erkannt - oft Ort der Gasentwicklung nicht präzise dargestellt - selten mehrere Bilder der Phänomene - große Unterschiede zwischen abstrakt und exakt - Legende fehlt oftmals	- Variable Wassermenge oftmals erkannt - Variable Verhältnis der Zugaben zwar erkannt, aber sehr unterschiedlich und teilweise nicht verständlich dargestellt - Kodierung nicht immer klar

Gründe für die Ausscheidung von Kontexten für die Pilot- und Hauptstudie waren eine schlechte Bandbreite, geringe Varianz oder kaum Unterschiede zwischen den Skizzen der Schülerinnen und Schüler. Ebenso führte eine ungenügende Kodierbarkeit, wie nicht genügend unterscheidbare Unterschiede in den Skizzen, und eine

schwierige Kodierbarkeit zu einer Ausscheidung. Eine umfangreiche Zusammenstellung der entstandenen Schwierigkeiten zeigt Tab. 19.

Eine Reduktion auf drei Kontexte wurde aufgrund der Studiengröße der geplanten Hauptstudie angestrebt. Der Kontext „Chromatographie“ bot wenig Unterschiede in den Skizzen und somit auch eine geringe Varianz. Des Weiteren wurden selten mehrere Bilder des Phänomens, sondern oftmals nur ein einzelnes Bild gezeichnet. Wenn eine Bildabfolge gezeichnet wurde, stellte diese vor allem Arbeitsschritte vor dem Phänomen, jedoch nicht des Phänomens selbst dar. Die Kodierbarkeit war gewährleistet.

Im Kontext „Kohlenstoffdioxid“ zeigte der Aufgabentyp „Beobachtung“ kaum Varianz. Die Skizzen waren sehr ähnlich, zudem wurde das Phänomen selten dargestellt. Beim Aufgabentyp „Experimentieren“ zwar die Variable „Wassermenge“ oft erkennbar. Die Kontrolle der zweiten Variablen (Menge der Zugabe von Stoffen) wurde jedoch sehr unterschiedlich, teilweise gar unerkennlich, dargestellt. Es zeigte sich ein Bodeneffekt, da die Skizzen selten den Anforderungen der höheren Niveaus entsprachen. Zudem war die Kodierbarkeit nicht immer gewährleistet.

Tab. 20 Items der Pilot- und Hauptstudie

Aufgabentyp Kontext	Beobachtung	Vergleich	Experimentieren
Brausetablette	Lösen einer Brausetablette in Wasser bei Raumtemperatur	Lösen einer Brausetablette in Mineralwasser bei Raumtemperatur	Untersuchung des Einflusses der Variablen Temperatur und Wassermenge auf das Lösen einer Brausetablette
Eis schmelzen	Schmelzen eines Eiswürfels bestreut mit Natriumchlorid	Schmelzen eines eingefärbten Eiswürfels in Salz- und Süßwasser	Untersuchung des Einflusses der Variablen Salzgehalt (Salz- und Süßwasser) und Wassermenge auf das Lösen eines Eiswürfels
Säuren und Basen	Reaktion von Essigsäure und Natronlauge mit Universalindikator	Reaktion von Essigsäure und Natronlauge bzw. Sodalösung mit Universalindikator	Untersuchung des Einflusses der Variablen Zugabe (Zitronensaft und Kernseife) und Teemengen auf die Farbe von Schwarztee.

Die Kontexte „Chromatographie“ und „Kohlenstoffdioxid“ schieden aus. Während die Skizzen des Aufgabentyps „Vergleich“ im Kontext „Kohlenstoffdioxid“ sehr va-

riabel und gut kodierbar waren, war die Kodierbarkeit im Kontext „Säuren und Basen“ kaum gewährleistet. Deshalb wurde der Kontext „Säuren und Basen“ durch den Kontext „Kohlensäure“ ersetzt. Dies war möglich, da bei der Erzeugung der Phänomene neben der Entwicklung von Kohlensäure auch Säure-Base-Indikatoren eingesetzt wurden. Die in der Studie eingesetzten Items sind in der Tab. 20 zusammengestellt.

Tab. 21 Abkürzung der Item-Bezeichnungen

Aufgabentyp Kontext	Beobachtung	Vergleich	Experimentieren
Brausetablette	BR_B	BR_V	BR_E
Eis schmelzen	EI_B	EI_V	EI_E
Säuren und Basen	SB_B	SB_V	SB_E

Im Folgenden werden für die Item-Bezeichnungen Abkürzungen verwendet. Die Kontexte werden mit zwei Großbuchstaben, die Aufgabentypen mit einem Großbuchstaben abgekürzt und beide mit einem Stich verbunden (siehe Tab. 21).

6.4.4 Pilotstudie

An der Pilotstudie (Juli - September 2019) nahmen 74 Schülerinnen und Schüler aus den Kantonen Zürich und Thurgau teil. Die Stichprobe bestand aus Schülerinnen und Schülern aus den Jahrgangsstufen 7 und 9 des Niveaus Sek A und Sek AB und der Jahrgangsstufe 9 des Gymnasiums.

Ziele der Pilotierung waren die Setzung der Progressionsniveaus, die Erstellung des Kodiermanuals mit Ankerbeispielen (siehe Anhang) sowie die damit verbundene Schulung der Kodierhelferinnen und -helfer. Das Aufgabenset bestand aus der Video-Vignette, dem Auftrag zur Erstellung der *wissenschaftlichen Skizzen* mittels Vignette und einem Set von Multiple-Select Fragen.

Der eingesetzte Kompetenztest war in einem Multi-Matrix-Design angelegt, sodass nicht jeder Schüler und jede Schülerin alle Items aus dem gesamten Kompetenztest beantworteten. Für die Datenerhebung der Pilotierung wurden drei Units zusammengestellt. Jede Unit bestand aus drei Items, wobei jeder Aufgabentyp und jeder der drei Kontexte berücksichtigt wurde (siehe Tab. 22). Die Schülerinnen und Schüler einer Klasse erhielten die gleichen zwei Units (siehe Tab. 23).

Tab. 22 Zusammensetzung Units Pilotstudie

Aufgabentyp	Beobachtung	Vergleich	Experimentieren
Unit 1	Brausetablette	Schmelzen	Säure und Basen
Unit 2	Schmelzen	Säure und Basen	Brausetablette
Unit 3	Säure und Basen	Brausetablette	Schmelzen

Nach einer fünfminütigen Einführung in die Regeln des *wissenschaftlichen Skizzierens* mit Hilfe des Theorieblattes (siehe Kapitel 6.3.3) wurden die Units bearbeitet, wobei das Vorgehen bei allen Units dasselbe war. Erst lasen die Schülerinnen und Schüler den Einführungstext, dann sahen sie die Video-Vignette das erste Mal im Plenum. Anschließend konnten sie die *wissenschaftliche Skizze* erstellen, sahen danach die Video-Vignette ein zweites Mal und bearbeiteten die Multiple-Select Fragen. Pro Item standen 10 Minuten zur Verfügung. Die Testung fand an einem Tag statt und umfasste total 65 Minuten. Durch das Befolgen des Erhebungsmanuals konnte die Durchführungsobjektivität gewährleistet werden.

Tab. 23 Rotationsschema Pilotstudie

Klasse 1	Units 1, 2
Klasse 2	Units 2, 3
Klasse 3	Units 1, 3
Klasse 4	Units 2, 3

Die in der Pilotierung erhobenen Daten zeigten nach der Evaluation der Resultate Skizzen auf allen Niveaus der Progression. Das Kodiermanual wurde mit Ankerbeispielen (siehe Anhang) aus der Pilotstudie erstellt und die Schulung der Kodierhelferinnen und -helfer erfolgte ebenfalls durch Skizzen aus der Pilotstudie. Weder die Items noch die Zusammenstellung der Units wurde für die Hauptstudie verändert.

6.4.5 Hauptstudie

An der Hauptstudie (Oktober 2019 - März 2020, Mai 2020) nahmen 374 Schülerinnen und Schüler aus den Kantonen Zürich und Thurgau teil. Die Stichprobe bestand aus Schülerinnen und Schülern aus der 7. und 9. Jahrgangsstufe des Niveaus Sek A, Sek AB, Sek B und der 9. Jahrgangsstufe Gymnasium.

Die Datenerhebung der Hauptstudie fand an einem Tag statt und umfasste eine Doppelschulstunde (90 Minuten). Die Datenerhebung erfolgte analog der Datenerhebung der Pilotstudie. Zu jedem Item des Kompetenztests beantworteten die Schülerinnen und Schüler ein Set an Multiple-Select Fragen. Zusätzlich wurden in der Hauptstudie als Begleitstudie die Kognitionsfähigkeit (KTF), die Lesefähigkeit (LGVT) und das Strategiewissen zum Experimentieren (NAW) erfasst (siehe Kapitel 6.3.5).

Alle 374 Schülerinnen und Schüler erhielten mit dem ersten Item einen demographischen Fragebogen (Geschlecht, Sprache(n)). 48.4 % der Lernenden waren weiblich. 40.9 % gaben an, zu Hause nur Deutsch/Schweizerdeutsch zu sprechen. 59.1 % aller Schülerinnen und Schüler gaben an, zuhause mehrsprachig aufzuwachsen. Bei 47.8 % wurde neben Deutsch eine weitere Sprache gesprochen. 11.3 % der Schülerinnen und Schüler sprachen mindestens zwei Fremdsprachen zuhause, 1.9 % machten keine Angaben. Die Schülerinnen und Schüler besuchten 24 verschiedene Klassen, geführt von 20 Lehrpersonen aus 13 unterschiedlichen Schulhäusern.

Tab. 24 Rotationsschema Hauptstudie

7. Klassen	9. Klassen	Units
Klassen 1 - 2	Klasse 13, 23	Units 1, 2
Klassen 3 - 4	Klasse 14, 20	Units 2, 3
Klassen 5 - 6	Klasse 16, 21	Units 1, 3
Klassen 7 - 8	Klasse 15, 22	Units 2, 1
Klassen 9 - 10	Klasse 17, 19	Units 3, 2
Klassen 11-12	Klasse 18, 24	Units 3, 1

Der eingesetzte Test zur Messung der Kompetenz des *wissenschaftlichen Skizzierens* war in einem Multi-Matrix-Design angelegt, so dass nicht jeder Schüler und jede Schülerin alle Items aus dem gesamten Kompetenztest beantworteten. Für die Datenerhebung der Hauptstudie wurden drei Units zusammengestellt. Jede Unit bestand aus einem Item jedes Aufgabentyps und aus einer Aufgabe aus jedem der drei Kontexte analog der Zusammenstellung der Units in der Pilotstudie (siehe Tab. 22). Jeder Lernende bearbeitete zwei Units, wobei die Schülerinnen und Schüler einer Klasse die gleichen Units erhielten (siehe Tab. 24). Die Durchführungsobjektivität wurde durch das Erhebungsmanual gewährleistet.

Reihenfolgeeffekte wurden durch die unterschiedliche Reihenfolge der Units verringert (Rost, 2004).

Bei den Klassen 19, 20 und 23 handelte es sich um Gymnasialklassen der 9. Jahrgangsstufe. Die Klassen 13, 17, 18 und 22 waren Sekundarstufe I Klassen der 9. Jahrgangsstufe des Niveaus A, die Klassen 14 und 24 waren Sekundarstufe I Klassen der 9. Jahrgangsstufe des Niveaus B und die Klassen 15, 16 und 21 wurden in der 9. Jahrgangsstufe auf dem Niveau AB unterrichtet. Die Klassen 21, 22 und 24 waren Sekundarstufe I Klassen der 9. Jahrgangsstufe. Die Klassen 1 - 12 waren Sekundarstufe I Klassen der 7. Jahrgangsstufe. Die Schülerinnen und Schüler der Klassen 1, 3, 7, 8, 9, 11 wurden auf dem Niveau A, die Lernenden der Klassen 2, 4, 6 auf dem Niveau B und die Schülerinnen und Schüler der Klassen 7, 10 und 12 auf dem Niveau AB unterrichtet.

6.5 Kodierung

6.5.1 Entwicklung des Kodiermanual

Ausgehend vom Progressionsmodell (siehe Kapitel 6.2) und den Items wurde das Kodiermanual mit Ankerbeispielen (siehe Anhang) entwickelt, in dem für alle drei Aufgabentypen die Grundanforderungen an das *wissenschaftliche Skizzieren* mit den Facetten „Eindeutigkeit“, „Idealisierung“, „Vollständigkeit“, „Abstraktion“ und „fachliche Korrektheit“ der Kategorie „*Wissenschaftlichkeit der Skizzen*“ (siehe Tab. 7) erläutert werden.

Die Aufgabentypen „Vergleich“ und „Experimentieren“ verfügen zusätzlich über eine aufgabentypspezifische Kategorie: „Unterschied erfasst und Vergleich dargestellt“ bzw. „Variablen erfasst und vollständig kontrolliert“.

Für die allgemeinen und die kontextspezifischen Objektmerkmale wurden Indikatoren (Beispiel siehe Anhang) entwickelt, die dichotom kodiert werden konnten (1=Kriterium erreicht, 0=Kriterium nicht erreicht). Die Kategorie „*Wissenschaftlichkeit der Skizzen*“ wird je nach Item durch elf bis vierzehn Indikatoren beschrieben. Die Kategorie „Unterschied erfasst und Vergleich dargestellt“ (Aufgabentyp „Vergleich“) umfasst je nach Kontext fünf bis sieben weitere Indikatoren, und die Kategorie „Variablen erfasst und vollständig kontrolliert“ (Aufgabentyp „Experimentieren“) wird durch jeweils sechs Indikatoren beschrieben, welche die Variablenkontrolle aufgreifen.

Die Entwicklung des Kodiermanuals basiert auf den *wissenschaftlichen* Skizzen der Schülerinnen und Schüler der Pilotstudie. Zu jedem Indikator wurden ein oder mehrere Ankerbeispiele (*wissenschaftliche Skizzen*) erstellt und mit einer genauen Beschreibung des Indikators ergänzt.

Die Überarbeitung und Verfeinerung des Kodiermanuals durch weitere Ankerbeispiele und Präzisierungen für die Beschreibung des Indikators erfolgte während der Schulung der studentischen Kodierhelferinnen und -helfer. Während der Hauptstudie wurde das Kodiermanual finalisiert, indem bis zum Ende der Interrater-Phase Ergänzungen in den Beschreibungen der Indikatoren und weitere Ankerbeispiele gemacht wurden. Ein Beispiel eines Kodiermanuals findet sich im Anhang.

6.5.2 Festlegung der Schwellenwerte

Die Schwellenwerte beschreiben die zu erreichenden Summen-Scores der erfüllten Indikatoren, damit ein nächsthöheres Niveau in der Kompetenz des *wissenschaftlichen Skizzierens* erreicht wird. Die Festlegung der Schwellenwerte erfolgte gestützt auf das Progressionsmodell (siehe Kapitel 6.2), indem Summen-Scores zum einen für die Indikatoren der einzelnen *Facetten* (siehe Tab. 7) der „*Wissenschaftlichkeit der Skizzen*“ und zum anderen für die aufgabentypspezifischen Kategorien (siehe Tab. 6) „Unterschied erfasst und Vergleich dargestellt“ und „Variablen erfasst und vollständig kontrolliert“ gebildet wurden.

Tab. 25 Überblick über die Begrifflichkeiten

Aufgabentyp	Kategorien		Facetten
Beobachtung	Wissenschaftlichkeit der Skizze (aufgeteilt in fünf Facetten)		Eindeutigkeit Idealisierung Vollständigkeit
Vergleich	Wissenschaftlichkeit der Skizze (aufgeteilt in fünf Facetten)	Unterschied erfasst und Vergleich dargestellt	Abstraktion Korrektheit
Experimentieren	Wissenschaftlichkeit der Skizze (aufgeteilt in fünf Facetten)	Variablen erfasst und vollständig kontrolliert	

Die Tab. 25 bietet einen Überblick über die Begrifflichkeiten: Aufgabentyp, Kategorie und *Facette*. Die *Indikatoren* (siehe Anhang) sind Bestandteil Kodiermanuals und werden dichotom mit 0/1 kodiert.

Bei der analytischen Niveaustufung durch Schwellenwerte von einer holistischen Niveaubeschreibung (siehe Kapitel 6.2) wurde auf ein Expertenrating verzichtet, da dieses die Grenzen des Projekts überschritten hätte. Die analytische Stufung muss jedoch begründet sein und den holistischen Beschreibungen entsprechen. Vorgegangen wurde mit einer Art „Partial Credit Model“, bei der die holistische Beschreibung und lineare Stufung der Anzahl erfüllter *Indikatoren* wie auch die natürliche Datenlage der Erfüllung der *Indikatoren* berücksichtigt wurde. Die natürliche Datenlage sind die Fähigkeiten der Lernenden dieser Studie in der Kompetenz des *wissenschaftlichen Skizzierens*. Dieses wird in Folgenden aufgezeigt und begründet. Für jede *Facette* wurde der Schwellenwert separat festgelegt, jedoch musste eine Schülerin oder ein Schüler die Schwellenwerte aller Facetten erfüllen, um das nächsthöhere Niveau zu erreichen (siehe bspw. Tab. 26). Zur Validierung dieser Festlegung wurde ermittelt, ob es Lernende gab, die, weil sie in einer *Facette* gegenüber den anderen Facetten deutlich tiefere Leistungen erreichten, in ein tieferes Niveau eingestuft wurden. Diese Überprüfung wird nachfolgend exemplarisch an Beispielen ausgeführt, indem das erreichte Niveau unter Ausschluss jeweils einer *Facette* ermittelt wurde.

Für die Festlegung der Schwellenwerte wurden folgende Regeln festgelegt: Für Items eines Aufgabentyps sollten für die Einheitlichkeit, wenn möglich, dieselben Schwellenwerte gelten, auch wenn die Anzahl *Indikatoren* bei einzelnen *Facetten* je nach Kontext leicht variierten. Für die *Facetten* der Kategorie „*Wissenschaftlichkeit der Skizze*“ sollten, wenn möglich, über die Aufgabentypen hinweg dieselben Schwellenwerte festgelegt werden. Für das Erreichen eines Niveaus muss mindestens eine gewisse Anzahl *Indikatoren* der *Facetten* erfüllt sein. Diese Werte sind in eckigen Klammern aufgeführt, wobei die erste Zahl für die erste *Facette* gilt, die zweite für die zweite und so fort (siehe Tab. 26). Bspw. muss für das Erreichen des *Niveaus 1* in der *Facette 1* kein *Indikator*, in der *Facette 2* zwei *Indikatoren* erfüllt sein. Für die *Facetten 3* und *4* muss mindestens einer der *Indikatoren* und für die *Facette 5* kein *Indikator* erfüllt sein.

Die Festlegung der Anzahl der als erreicht bewerteten *Indikatoren* und damit der Gewichtung der *Facetten* innerhalb der Niveaus erfolgte auf der Basis der holistischen Beschreibungen des Progressionsmodells (siehe Kapitel 6.2) und wird durch die natürliche Datenlage gestützt.

Tab. 26 Schwellenwerte: Kontexte „Brausetablette und „Säuren und Basen“, Aufgabentyp „Beobachtung“

Übergang Niveau	Wissenschaftlich- keit der Skizze	Wissenschaftlich- keit der Skizze	Facette 1	Eindeutigkeit
Kontext	„Brausetablette „Säuren und Basen“	„Eis schmelzen“	Facette 2	Idealisierung
2 → 3	[≥2, ≥3, ≥3, ≥2, ≥1]	[≥2, ≥3, ≥3, ≥2, ≥1]	Facette 3	Vollständigkeit
1 → 2	[≥1, ≥3, ≥2, ≥2, ≥0]	[≥1, ≥3, ≥2, ≥2, ≥0]	Facette 4	Abstraktion
0 → 1	[≥0, ≥2, ≥1, ≥1, ≥0]	[≥0, ≥2, ≥1, ≥1, ≥0]	Facette 5	fachliche Korrektheit

Zur analytischen Schwellenwertbestimmung anhand von dichotomen *Indikatoren* zur Abbildung einer holistischen Niveaubeschreibung wurden zwei Faktoren berücksichtigt: Erster Faktor ist die Anzahl erfüllter *Indikatoren* pro *Facette* zur Erreichung eines bestimmten Niveaus, und der zweite Faktor ist die Gewichtung der einzelnen *Facetten* bezüglich der Niveaus. Durch Berücksichtigung dieser beider Faktoren kann die holistische Beschreibung des Kompetenzstrukturmodells durch die dichotome Kodierung umgesetzt werden. Der erste Faktor wird am Beispiel der *Facette 1* (Eindeutigkeit) betrachtet. Die *Facette 1* wird wie folgt beschrieben (siehe Tab. 7): Eindeutigkeit (eindeutig interpretierbar): relevante Objektmerkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze eindeutig interpretierbar. Für die Erreichung des *Niveaus 3* müssen die relevanten Objektmerkmale „dargestellt“ und die Skizze „eindeutig“ sein. Für die *Niveaus 2* bis *0* ist der Erreichungsgrad in der holistischen Beschreibung niedriger: *Niveau 2* „Objektmerkmale mindestens dargestellt/ Skizze größtenteils eindeutig“; *Niveau 1* „Objektmerkmale mindestens größtenteils dargestellt/ Skizze mindestens teilweise eindeutig“ und *Niveau 0* „Objektmerkmale nicht oder ansatzweise dargestellt/ Skizze nicht oder ansatzweise eindeutig“. Die Festlegung der Anzahl *Indikatoren*, die für das Erfüllen der *Facette 1* für ein bestimmtes Niveau erreicht werden müssen, (siehe Tab. 26) erfolgte analytisch, indem der Erfüllungsgrad den holistischen Beschreibungen (siehe Tab. 9 - 11) linear abgepasst wurde. Für das Erreichen des *Niveaus 1* wird in dieser *Facette* noch kein erfüllter *Indikator* erwartet; dies zum einen, weil die holistischen Beschreibungen für das *Niveau 1* sehr schwach formuliert sind, zum anderen - und dies ist bedeutsamer -, weil in dieser *Facette* gesamthaft nur zwei *Indikatoren* vorhanden sind. Für die Festlegung, welche *Facetten* bereits teilweise für das Erreichen des *Niveaus 1* erfüllt werden müssen

(siehe Tab. 26) wurden sowohl die gesamte Anzahl *Indikatoren* pro *Facette* als auch die holistischen Beschreibungen (siehe Tab. 9 - 11) berücksichtigt. Die holistische Beschreibung der *Facette 2* für das *Niveau 1* fordert bspw. bereits „mindestens teilweise Irrelevantes, teilweise unfokussiert, falls Text vorhanden, sagen Text und Skizze dasselbe (redundant)“ wohingegen für die *Facette 5* das *Niveau 1* noch „mindestens größtenteils fachlich nicht korrekt (fehlerhaft, Fehlendes)“ sein darf. Für das Erfüllen der *Facette 5* im *Niveau 1* muss daher kein *Indikator* erfüllt sein.

Diese analytische Schwellenwertbestimmung anhand von dichotomen *Indikatoren* zur Abb. einer holistischen Niveaubeschreibung wird nachfolgend an zwei Beispielen (Items) bezüglich der natürlichen Datenlage aufgezeigt. Diese Items sind: BR_B (Beobachtung: Brausetablette) und EI_V (Vergleich: Eis schmelzen):

Zur Abbildung der natürlichen Datenlage der Studie wurden die Summe der erfüllten *Indikatoren* je *Facette* und deren gewichtete Prozente ermittelt (siehe Tab. 27 und 28).

Tab. 27 gewichtete Prozente der Erfüllung der Facette und Rang BR_B

Facette	1	2	3	4	5
gewichtete Prozente	64 %	90 %	51 %	46 %	38 %
Rang	2	1	3	4	5

Gewichtete Prozente wurden mit dem Grundwert „*erreichbare Indikatoren*“ und dem Prozentwert „*erfüllte Indikatoren*“ je *Facette* ermittelt. Daraus wurde der Rang der Erfüllung der *Facette* ermittelt. Den ersten Rang erhielt diejenige *Facette*, die am besten erfüllt wurde (siehe Tab. 27 und 28).

Tab. 28 gewichtete Prozente der Erfüllung der Facette und Rang EI_V

Facette	1	2	3	4	5
gewichtete Prozente	83 %	91 %	61 %	77 %	58 %
Rang	2	1	4	3	5

Für das Item BR_B (Beobachtung: Brausetablette) ergeben sich folgende Werte (siehe Tab. 27): Die Reihenfolge der Facetten zeigt an, dass die *Facette 5* (fachliche Korrektheit) am schwierigsten für die Lernenden ist. Die *Facette 2* (Idealisierung) und die *Facette 1* (Eindeutigkeit) waren für die Lernenden der Stichprobe am einfachsten zu bewältigen. Vergleicht man diese Reihenfolge der Erfüllung der Facetten mit der

Festlegung der Schwellenwerte für dieses Item (siehe Tab. 26) zeigt sich dieselbe Reihenfolge. Die *Facette 2* muss bspw. bereits für das Erfüllen den *Niveaus 2* (siehe Tab. 26) vollständig erfüllt sein. Die *Facette 5* hingegen muss erst für das Erfüllen des *Niveaus 3* teilweise erfüllt sein. Die natürliche Datenlage wird in der Festlegung der Schwellenwerte abgebildet. Dasselbe Bild zeigt auch das Item EI_V (Vergleich: Eis schmelzen) (siehe Tab. 28 und 31):

Es wurde für die Items des Kontexts Brausetablette und für das Item EI_V auch eine Prüfung des Niveaus gemacht, indem das Niveau erneut auf der Basis der festgelegten Schwellenwerte berechnet wurde, aber jeweils eine *Facette* für diese Berechnung nicht berücksichtigt wurde. Aufgeführt wurde die Summe der Abweichungspunkte der Niveaus (siehe Tab. 29) als Summe der 374 Teilnehmenden der Studie. Die Abweichungspunkte umfassen die Summe der Abweichungen: Das heißt wenn ein Schüler oder eine Schülerin ohne Berücksichtigung der *Facette 3* das *Niveau 3* erreicht, mit der *Facette 3* jedoch nur das *Niveau 1*, sind das 2 Abweichungspunkte.

Tab. 29 Abweichungspunkte des Niveaus unter Ausschluss einer Facette des Kontext Brausetablette und des Items EI_V

Niveau bei Nichtberücksichtigung von Facette	1	2	3	4	5
Abweichungspunkte (BR_B)	0	65	174	19	8
Abweichungspunkte (BR_V)	1	30	273	17	2
Abweichungspunkte (BR_E)	20	42	64	25	3
Abweichungspunkte (EI_V)	0	46	97	14	17

Die Vollständigkeit (*Facette 3*) hat starkes Gewicht auf die Niveaus und schränkt ein, auf ein höheres Niveau zu kommen (siehe Tab. 26, 30 - 33). Das Progressionsmodell (siehe Kapitel 6.2) beschreibt diese *Facette* wie folgt: „Vollständigkeit: Vorgänge /Abläufe dargestellt (Bildergeschichte/Nummern/ Verwendung von Pfeilen, um Vorgänge zu zeigen)“. Für das Erreichen des *Niveaus 3* müssen diese Vorgänge/Abläufe „vollständig“, für das *Niveau 2* „mindestens größtenteils“, für das *Niveau 1* „mindestens teilweise“ und für das *Niveau 0* „nicht oder nur teilweise“ dargestellt sein. Ausgehend vom Progressionsmodell (siehe Kapitel 6.2) wurden Schwellenwerte für diese *Facette* (siehe Tab. 26, 29 - 33) gesetzt: Die *Facette 3* verfügt über drei *Indikatoren*; davon muss einer bereits für die Erreichung des *Niveau 1* erfüllt sein (siehe Tab. 26, 30- 33).

Im Item BR_B (siehe Tab. 26 und 29) erfüllen jedoch 47 Lernende keinen dieser drei *Indikatoren* und erreichen dadurch nur das *Niveau 0* (siehe Tab. 26). Weitere 65 bzw. 57 Lernende erreichen nur 1 bzw. 2 *Indikatoren*. Das heißt, sie erreichen dadurch das *Niveau 1* bzw. 2. Auch für die anderen Items sieht man, dass die *Facette 3* (Vollständigkeit) und die gesetzten Schwellenwerte die gesamte Bewertung dominieren.

Die Vollständigkeit ist inhaltlich äußerst relevant für die Kompetenz des *wissenschaftlichen Skizzierens*, wie die Validierung der Qualitätsmerkmale des Kompetenzstrukturmodells (siehe Kapitel 6.2) zeigt. Das teilweise bzw. vollständige (je nach Niveau) Erfüllen der *Facette* „Vollständigkeit“ ist zentral für die Kompetenz des *wissenschaftlichen Skizzierens*, denn dadurch, dass Vorgang und Ablauf vollständig in einer Skizze dargestellt werden, kann eine Skizze eine förderliche Basis für die Erkenntnisgewinnung darstellen (siehe Tab. 3 und Kapitel 3.2). Bei der Betrachtung der Abweichungen in den anderen Facetten (siehe Tab. 29) fällt auf, dass die *Facette 1* gar keine, die *Facetten 4* und 5 kaum Niveau-Abweichungen bei Nichtberücksichtigung dieser jeweiligen *Facette* zeigen. Die *Facette 2* zeigt Abweichungspunkte (siehe Tab. 29), die vor allem darauf zurückzuführen sind, dass diese *Facette* bereits für das *Niveau 2* vollständig erfüllt sein muss (siehe Tab. 26, 30 - 33). Die Dominanz der *Facette* Vollständigkeit über die anderen Facetten bedeutet für die Validierung des Tests Einschränkungen. Der Test ist so nicht in der Lage, die Dominanz dieser *Facette* wiederzugeben, wodurch Lernende möglicherweise in ein tieferes Niveau in Bezug auf ihre Fähigkeiten anderer *Facetten* eingeordnet werden. Eine mögliche Anpassung wäre eine größere Anzahl an *Indikatoren* in dieser *Facette*, um die Beschreibung feinschichtiger zu gestalten und so die Dominanz dieser *Facette* über die anderen zu verringern.

Nachfolgend wird die Festlegung der Schwellenwerte getrennt nach Aufgabentyp und Kontext einzeln aufgezeigt und die unterschiedliche Anzahl *Indikatoren* der Schwellenwerte begründet. Aufgrund der holistischen Niveaubeschreibungen im Progressionsmodell (siehe Kapitel 6.2) sowie der Abbildung der natürlichen Datenlage (siehe Tab. 27 und 28) wurde die Anzahl in der *wissenschaftlichen Skizze* erkannter *Indikatoren* pro *Facette* festgelegt. Die Anzahl zu erfüllender *Indikatoren* pro Schwellenwert weicht beim Aufgabentyp „Beobachtung“ nur im Kontext „Eis schmelzen“ ab.

Tab. 26 zeigt die Schwellenwerte für die Übergänge der Niveaus des Aufgabentyps „Beobachtung“. Beim Kontext „Eis schmelzen“ liegt für die *Facette* „Abstraktion“

nur ein *Indikator* vor. Das Progressionsmodell legt fest, dass ab dem *Niveau 1* diese *Facette* mehrheitlich erfüllt ist, daher wurde der Schwellenwert auch für den Kontext „Eis schmelzen“ auf einen *Indikator* festgelegt, obwohl es sich dabei um den einzigen *Indikator* dieser *Facette* handelt. In der *Facette 5* wurde wegen nicht idealer Interrater-Übereinstimmung ein *Indikator* gestrichen. Dadurch ist in dieser *Facette* für diesen Kontext nur ein Indikator vorhanden. Die Erfüllung dieses *Indicators* ist erforderlich, um das *Niveau 3* in dieser *Facette* zu erreichen. Bei den anderen beiden Kontexten wird für die Erreichung des *Niveaus 3* nur die Erfüllung einer der beiden *Indikatoren* vorausgesetzt. Das Progressionsmodell setzt für diese *Facette* eine „größtenteils fachliche Korrektheit“ voraus.

Aufgrund der leicht unterschiedlichen Anzahl Indikatoren im Aufgabentyp „Vergleich“ variiert die Anzahl zu erreichender *Indikatoren* bei den Schwellenwerten geringfügig. Die Abweichung ist kontextbedingt. Bspw. weisen die beiden Phänomene im Kontext „Eis schmelzen“ mehr beobachtbare Unterschiede auf als die Phänomene im Kontext „Säuren und Basen“.

Tab. 30 Schwellenwerte: Kontext „Brausetablette“, Aufgabentyp „Vergleich“

Übergang Niveau	Wissenschaftlichkeit der Skizze	Unterschied erfasst und Vergleich dargestellt
2 → 3	[≥2, ≥3, ≥3, ≥2, ≥1]	[≥5]
1 → 2	[≥1, ≥2, ≥2, ≥2, ≥0]	[≥4]
0 → 1	[≥0, ≥2, ≥1, ≥1, ≥0]	[≥2]

Die Schwellenwerte sind, bis auf die *Facette 2* (Idealisierung), bei allen drei Kontexten identisch zu den Schwellenwerten des Aufgabentyps „Beobachtung“, ebenso die *Facette 5* (Korrektheit) beim Kontext „Eis schmelzen“ (siehe Tab. 30 - 32). Die Abweichung in der *Facette 2* kommt dadurch zustande, dass im Bereich der Idealisierung bei der Darstellung von zwei Phänomenen für die Beschreibung des *Niveaus 2* die Idealisierung nicht vollständig umgesetzt werden muss.

Tab. 31 Schwellenwerte: Kontext „Eis schmelzen“, Aufgabentyp „Vergleich“

Übergang Niveau	Wissenschaftlichkeit der Skizze	Unterschied erfasst und Vergleich dargestellt
2 → 3	[≥2, ≥3, ≥3, ≥2, ≥3]	[≥6]
1 → 2	[≥1, ≥2, ≥2, ≥2, ≥1]	[≥5]
0 → 1	[≥0, ≥2, ≥1, ≥1, ≥0]	[≥3]

Tab. 32 Schwellenwerte: Kontext „Säuren und Basen“, Aufgabentyp „Vergleich“

Übergang Niveau	Wissenschaftlichkeit der Skizze	Unterschied erfasst und Vergleich dargestellt
2 → 3	[≥2, ≥3, ≥3, ≥2, ≥1]	[≥5]
1 → 2	[≥1, ≥2, ≥2, ≥1, ≥0]	[≥3]
0 → 1	[≥0, ≥2, ≥1, ≥1, ≥0]	[≥1]

Die Abweichung in der *Facette 5* (fachliche Korrektheit) für den Kontext „Eis schmelzen“ ist dadurch bedingt, dass bei diesem Kontext einer der *Indikatoren* in zwei Teilbereiche aufgeteilt wurde, wobei der eine nur die Darstellung eines Eiswürfels in der Skizze abfragt. Die Erhöhung der Anzahl korrekter *Indikatoren* für das *Niveau 2* und das *Niveau 3* entspricht der holistischen Beschreibung dieser *Facette* im Progressionsmodell wie auch der Anzahl *Indikatoren* dieser *Facette*. Bei der Festlegung der Anzahl *Indikatoren* für die Kategorie „Unterschied erfasst und Vergleich dargestellt“ wurde die Beschreibung der Niveaus im Progressionsmodell wie auch die jeweiligen *Indikatoren* der Kontexte berücksichtigt. Da nicht alle *Indikatoren* dasselbe abfragen, variiert auch die Anzahl *Indikatoren* pro Niveau.

Tab. 33 Schwellenwerte: Aufgabentyp „Experimentieren“, alle drei Kontexte

Übergang Niveau	Wissenschaftlichkeit der Skizze	Variablen erfasst und vollständig kontrolliert
2 → 3	[≥1, ≥3, ≥3, ≥3, ≥3]	[≥6]
1 → 2	[≥1, ≥3, ≥2, ≥2, ≥2]	[≥4]
0 → 1	[≥0, ≥2, ≥1, ≥1, ≥0]	[≥3]

Die Tab. 33 zeigt die Schwellenwerte für den Aufgabentyp „Experimentieren“. Da bei diesem Aufgabentyp weder eine unterschiedliche Anzahl *Indikatoren* noch inhaltliche Unterschiede in der Bedeutung der *Indikatoren* in Abhängigkeit der Kontexte vorliegen, konnten die Schwellenwerte für den ganzen Aufgabentyp identisch festgelegt werden. Eine Abweichung der Anzahl *Indikatoren* pro Niveau von den anderen beiden Aufgabentypen im Bereich der „*Wissenschaftlichkeit der Skizze*“ hängt einerseits von den Abweichungen der holistischen Beschreibungen im Progressionsmodell (siehe Kapitel 6.2), andererseits von der unterschiedlichen Anzahl der *Indikatoren* je *Facette* zwischen den Aufgabentypen ab. Bspw. wird die *Facette 1* (Eindeutigkeit) für den Aufgabentyp „Experimentieren“ nur durch einen *Indikator* gegeben. Die Anzahl *Indikatoren* wird auch in der Kategorie „Variablen erfasst und vollständig

kontrolliert“ durch die Beschreibung der Niveaus im Progressionsmodell und durch die Inhalte der *Indikatoren* berücksichtigt. Für das Erreichen des *Niveaus* 3 (höchstes Niveau) wird die Erfüllung aller *Indikatoren* vorausgesetzt.

6.5.3 Aufgabentypabhängige Unterschiede der Items und der Indikatoren

Tab. 34 Gemeinsamkeiten und Unterschiede zwischen den Aufgabentypen Beobachtung/ Vergleich und Experimentieren

Facette	Gemeinsamkeiten	Unterschiede	
	zwischen den Aufgabentypen „Experimentieren“ und „Vergleich“/„Beobachtung“	„Vergleich“/ „Beobachtung“	„Experimentieren“
Eindeutigkeit	Beschreibung Kompetenzstrukturmodell (KSM) identisch Indikator: „ist alles dargestellt“	2-3 Indikatoren zusätzlich wesentliche Objektmerkmale der Reaktion dargestellt	1 Indikator
Idealisierung	Schwellenwerte identisch Beschreibung KSM identisch		
Vollständigkeit	Schwellenwerte nahezu identisch	mehrere Bilder der Reaktion dargestellt	Variablenkontrolle dargestellt Phänomen nicht dargestellt
Abstraktion	Beschreibung KSM identisch Indikator zur generellen Abstraktion	1 Indikator Abstraktion: Objektmerkmale	2 Indikatoren Abstraktion: Variablen
fachliche Korrektheit	Beschreibung KSM nahezu identisch	Korrektheit der wesentlichen Objektmerkmale	4 Bechergläser dargestellt / Korrektheit der Variablenkontrolle
Aufgabentyp-spezifische Kategorie		Unterschied erfasst und Vergleich dargestellt	Variablen erkannt und Variablenkontrolle dargestellt

Das Kompetenzstrukturmodell ist für die Kategorie „*Wissenschaftlichkeit der Skizzen*“ der Aufgabentypen „Beobachtung“ und „Vergleich“ identisch. Der Aufgabentyp „Vergleich“ verfügt über eine zusätzliche Kategorie (Unterschied erfasst und

Vergleich dargestellt). Der Aufgabentyp „Experimentieren“ unterscheidet sich in der Kategorie „*Wissenschaftlichkeit der Skizzen*“ teilweise von den beiden anderen Aufgabentypen. Tab. 34 zeigt Unterschiede und Gemeinsamkeiten auf.

6.5.4 Kontextabhängige Unterschiede der Items und der Indikatoren

Der Unterschied zwischen den Aufgabentypen „Beobachtung“ und „Vergleich“ ist dadurch bedingt, dass beim Aufgabentyp „Vergleich“ zwei Phänomene beobachtet werden. Für den Aufgabentyp „Vergleich“ wird in den Facetten „Eindeutigkeit“, „Vollständigkeit“ und auch „fachliche Korrektheit“ deshalb verlangt, dass beide Phänomene eindeutig, vollständig und korrekt dargestellt werden, im Gegensatz zum Aufgabentyp „Beobachtung“, bei dem nur ein Phänomen vollständig und korrekt dargestellt werden muss. Diese Unterschiede zeigen sich im Kodiermanual und haben somit Einfluss auf die Kodierung.

Tab. 35 Objektmerkmale, die zu Indikatoren der Facette „fachliche Korrektheit“ führen

Aufgaben	Merkmalsbeschreibung
BR_B	- Blasen gehen von Brausetablette aus - Bewegung der Brausetablette
EI_B	- Furchen (gestrichener Indikator) - Schmelzvorgang mit Salz rascher
SB_B	- Schlieren-Bildung - Farbwechsel
BR_V	- Blasen gehen von Smartie/Brausetablette aus - Blasengröße - Ausbreitungsart der Blasen
EI_V	- Einfärbung Wasser korrekt - korrekter Schmelzvorgang (bei Salzwasser langsamer) - Eiswürfel eingezeichnet - Position Eiswürfel
SB_V	- Farbwechsel zu einem Regenbogen - Blasen entstehen am korrekten Ort

Kontextabhängige Unterschiede der Items und der Indikatoren haben Einfluss auf die Itemschwierigkeit. In der Facette „Eindeutigkeit“ wird im Kontext „Säuren und Basen“ für alle drei Aufgabentypen zusätzlich zur *wissenschaftlichen Skizze* entweder eine Legende oder eine Beschriftung direkt in der *wissenschaftlichen Skizze* zur Erfüllung des Indikators verlangt. Für die anderen beiden Kontexte wurde darauf ver-

zichtet, da diese auch ohne Legende oder Beschriftung die Grundanforderung „Eindeutigkeit“ erfüllen (siehe Kapitel 3.5.2 und 6.2). In der Facette „Idealisierung“ skizzieren die Lernenden bei den Aufgabentypen „Beobachtung“ und „Vergleich“ im Kontext „Säuren und Basen“ oftmals zusätzlich einen Reagenzglas-Ständer. Dies stellt eine Fehlerquelle dieses Kontexts dar.

Bei der Facette „Vollständigkeit“ sind zwei Indikatoren über alle Items hinweg identisch, ein weiterer Indikator bzw. zwei weitere Indikatoren sind abhängig von aufgabenspezifischen Objektmerkmalen. Die Items des Aufgabentyps „Vergleich“ des Kontextes „Eis schmelzen“ verfügen über zwei weitere Indikatoren. Für das Erreichen des Niveaus 3 ist jedoch neben den beiden allgemeinen Indikatoren nur die Erfüllung eines weiteren notwendig. Die Facette „Abstraktion“ ist für alle Items bis auf das Item des Aufgabentyps „Vergleich“ des Kontextes „Säuren und Basen“ identisch. Dieses Item verfügt im Gegensatz zu allen anderen über drei Indikatoren für diese Facette. Das Niveau 3 wird erreicht, wenn zwei von drei Indikatoren erfüllt sind. Die Facette „fachliche Korrektheit“ weist am meisten kontextabhängige Unterschiede auf, da diese Facette die korrekte Darstellung der relevanten Objektmerkmale in den Indikatoren abfragt. Obwohl die Anzahl Indikatoren variiert, ist für die Erreichung des Niveaus 3 jeweils nur die Erfüllung eines Indikators notwendig. Eine Ausnahme bildet beim Aufgabentyp „Vergleich“ der Kontext „Eis schmelzen“. Hierfür ist die korrekte Darstellung von drei Indikatoren notwendig, um das Niveau 3 zu erreichen. In der Tab. 35 sind alle Objektmerkmale, die zu Indikatoren der Facette „fachliche Korrektheit“ führen, aufgeführt.

Der Kontext „Säuren und Basen“ ist aufgrund von kontextabhängigen Objektmerkmalen, wie Farbe, Notwendigkeit der Beschriftung und Vorhanden-Sein von Reagenzglas-Ständern, in der Video-Vignette anspruchsvoller als die anderen beiden Kontexte.

6.6 Reliabilität der Kodierung

Die Interrater-Reliabilität (Interrater-Übereinstimmung) ist ein Maß für die Personenabhängigkeit der Kodierung eines Tests (Wirtz & Caspar, 2002) und damit für die Auswertungsobjektivität. Es gibt im Wesentlichen zwei Ursachen für mangelnde Übereinstimmung zwischen Interratern. Einerseits sind sich die Rater nicht einig darüber, welches Merkmal beurteilt werden soll, insbesondere welche Aspekte des

Merkmals berücksichtigt werden müssen, welche Gewichtung diese haben und wie sie sich zum Gesamturteil zusammenfügen. Andererseits können die Rater auch zu unterschiedlichen Urteilen kommen, weil sie die Ausprägung der zu beurteilenden Merkmalsdimension unterschiedlich einschätzen (Wirtz & Caspar, 2002). Letzteres kommt gerade bei der Kodierung von Skizzen häufig vor. Bei Multiple-Choice Aufgaben liegt die Interrater-Reliabilität in der Regel bei nahezu 100 %, da Abweichungen in der Kodierung nur von Flüchtigkeitsfehlern oder der unterschiedlichen Beurteilung unerwarteten Antwortverhaltens abhängig sind.

Tab. 36 Prozentsatz verglichener Testhefte

	Anteil doppelt kodierter Testhefte
Unit 1	40 von 246 16.3 %
Unit 2	40 von 256 15.6 %
Unit 3	42 von 245 17.1 %

Durchschnittlich 16 % der Schülerskizzen wurden im Laufe der Hauptstudie von zwei Personen bewertet, wobei jeweils eine Person sämtliche Skizzen eines Aufgabentyps bewertete (Masterspur). An eine erste Schulung mit je fünf Skizzen pro Kontext und Aufgabentyp in der Pilotierungsphase schloss sich eine gemeinsame Trainingsphase mit 18 % der Skizzen an, in der allfällige Kodierungsunterschiede besprochen wurden und auch Ergänzungen im Kodiermanual erfolgten. Für die Berechnung der Interrater-Reliabilität wurden rund 16 % aller Testhefte doppelt kodiert und die Übereinstimmung der Kodierungen in den einzelnen Kriterien aus dem Kodiermanual verglichen (siehe Tab. 36). Eine vollständige Übersicht der Interrater-Auswertung (Übereinstimmung/Cohens Kappa/Gwet AC1) sowie von Interferenz, Schiefe, Mittelwert und Standardabweichung ist im Anhang aufgeführt. Pro Item wurden zwischen elf bis einundzwanzig Indikatoren bewertet.

6.6.1 Prozentuale Übereinstimmung

Die prozentuale Übereinstimmung ist das einfachste Maß der Auswertungsobjektivität. Sie gibt an, wie viele Übereinstimmungen ein oder mehrere Rater haben (Fleiss & Cohen, 1973). Die prozentuale Übereinstimmung hat als Maß der Übereinstimmung zwischen zwei Ratern nur eine beschränkte Aussagekraft, da Zufallsübereinstimmungen nicht berücksichtigt werden (Wirtz & Caspar, 2002). Als Folge überschätzt die

prozentuale Übereinstimmung die „wahre“ Übereinstimmung. Grundsätzlich gilt eine Übereinstimmung von mindestens 80 % als Voraussetzung für eine zufriedenstellende Interrater-Reliabilität (Wirtz & Caspar, 2002).

6.6.2 Cohens Kappa

Cohens κ (Kappa) wurde von Cohens (1960) durch eine Modifikation des von Scott (1955) entwickelten π -Koeffizienten weiterentwickelt. Es basiert auf der prozentualen Übereinstimmung, verfügt aber über eine Korrektur der durch den Zufall erwarteten Übereinstimmung und liefert eine standardisierte Maßzahl zwischen -1 und 1 als Resultat, wobei der Wert 1 von Cohens κ perfekte Übereinstimmung und der Wert -1 vollkommen unterschiedliche Einschätzung der beiden Rater bedeutet. Die Werte 1 und -1 werden nur erreicht, wenn die Randsummenverteilung der beiden Rater gleich ist, was nichts anderes bedeutet, als dass die Rater die verschiedenen Werte (Bewertungen) gleich oft verteilt haben. Eine positive Zahl zeigt dabei eine über dem Zufall liegende Übereinstimmung an (Wirtz & Caspar, 2002).

Allgemein gilt für einen Indikator ein Wert von $\kappa > 0.75$ als sehr gute, ein Wert für κ zwischen 0.61 und 0.75 als gute Übereinstimmung (Fleiss & Cohen, 1973; Wirtz & Caspar, 2002). Werte für κ , die zwischen 0.40 und 0.60 liegen, können als akzeptabel gewertet werden. Je nach Literatur werden verschiedene Gültigkeitsbereiche diskutiert (Wirtz & Caspar, 2002). Festzuhalten ist, dass je nach den zu bewertenden Objekten und Merkmalen ein Kappa von 0.50 als zufriedenstellend (asymmetrische Daten und offene Antwort-Formate) oder aber ein Kappa von 0.80 als zu niedriger Wert (bspw. Multiple Choice Test) angesehen werden kann. Bei einer sehr asymmetrischen Verteilung der Werte, das bedeutet, wenn bei einer dichotomen Bewertung vor allem ein Wert vorliegt, der andere aber kaum geratet wurde, kann auch bei einer guten prozentualen Übereinstimmung der Indikator κ schlecht ausfallen dies vor allem, wenn beim κ -Koeffizienten die Schiefe (sehr asymmetrische Verteilung) der gerateten Daten nicht berücksichtigt wird (Gwet, 2002).

6.6.3 Gwet's AC1

Der Koeffizient Cohens κ ersagt bei einer sehr asymmetrischen Verteilung der gerateten Werte, vor allem, wenn die Randsummenverteilung deutlich von 1 abweicht. Gwet (2002) führte eine neue Statistik ein, die deutlich robuster bezüglich asymmetrisch verteilter Daten ist, und verwendet den Koeffizienten Gwet AC1 als Alternative

zu Cohens κ . Die Gwet AC1 Statistik vermeidet bei ungleich hohen Verteilungen die Probleme von Cohens κ . Für den Fall der perfekten Übereinstimmung beider Rater nimmt Gwet AC1 den Wert 1 an (Gwet, 2002). Für den Fall gleichverteilter Bewertungen sind Cohens κ und Gwet AC1 also identisch. Die Gütekriterien für die Cohens κ -Werte gelten laut Schori (2006) auch für die Werte von Gwet AC1. Im Bereich von 0.81 - 1.00 gilt die Übereinstimmung als sehr hoch, im Bereich von 0.61 - 0.81 als hoch.

6.6.4 Interrater-Reliabilität

Eine vollständige Übersicht der Interrater Auswertung (Übereinstimmung/Cohens /Kappa/Gwet AC1) inklusive Interferenz, Schiefe, Mittelwert und Standardabweichung ist im Anhang aufgeführt. Die prozentualen Übereinstimmungen der Kodierungen (Interrater-Reliabilität) lagen bei allen verglichenen Indikatoren (153 Indikatoren) zwischen 83.3 % und 100 %.

Vierzig Indikatoren (26.3 %) zeigen eine Übereinstimmung von 100 %. Dies kann zu einem „nicht definierten“ Cohens κ (Cohen, 1960) führen, falls die Schülerinnen und Schüler, deren Skizzen bei diesem Indikator doppelt kodiert wurden, alle entweder das Kriterium erfüllen oder alle das Kriterium nicht erreichen. Weiter kann für den Fall der 100 % Interrater-Übereinstimmung das Cohens κ (Cohen, 1960) $\kappa = 1$ sein. Für beide Fälle ($\kappa = 1$ und $\kappa =$ nicht definiert) nimmt das Gwet AC1 (Gwet, 2008) einen Wert von 1 an. Die Mittelwerte des Erfüllens dieser Indikatoren aller Lernenden weichen von der totalen Übereinstimmung (100 % oder 0 %) ab.

Tab. 37 Zusammenstellung niedrig-inferenter Indikatoren

Aufgabentyp	Kontext	Facette	Indikator
Beobachtung	Brause-tablette	Eindeutigkeit	Ist der Vorgang des Experiments nur mit Legende erkennbar und nachvollziehbar? (Wasser, Brausetablette, Brausetablette im Wasser)
Vergleich	Eis schmelzen	Vergleich erfasst und Unterschied dargestellt	Sind die zu vergleichenden Sachverhalte nebeneinander dargestellt? (Salz- und Süßwasser?)
Zusammenhang	alle drei Kontexte	Fachliche Korrektheit	Wurden 4 Becher dargestellt?

Hundertprozentige Übereinstimmungen beim Rating gibt es bei Indikatoren, die eindeutig formuliert sind und deren *wissenschaftliche Skizzen* unabhängig von der Interpretation sind. Beispiele solcher als „niedrig-inferent“ geltenden Indikatoren sind in Tab. 37 aufgeführt. Die Inferenz wurde mittels Konsensbildung der verschiedenen Kodierer ermittelt.

Ein weiterer Grund für eine hundertprozentige Übereinstimmung liegt im Boden- und Deckeneffekt (Rost, 2004) einzelner Indikatoren. Einen Bodeneffekt zeigt der Indikator „Sind die wesentlichen Materialien dargestellt: Becher, Wasser und Brausetablette?“ der Facette „Idealisierung“ im Kontext „Brausetablette“ im Aufgabentyp „Beobachtung“. Der Mittelwert von 0.963 (siehe Anhang) zeigt an, dass fast alle Lernenden diesen Indikator erfüllt haben. Die Übereinstimmung von 100 % wie auch ein Cohens Kappa von 1 ($\kappa=1$), lassen sich dadurch erklären, dass nahezu alle Lernenden diesen Indikator korrekt erfüllt haben. Dieser Indikator ist inferent.

Einen Deckeneffekt zeigt der Indikator „Ragt der Eiskwürfel beim Salzwasser mehr aus dem Wasser?“ der Facette „fachliche Korrektheit“ im Kontext „Eis schmelzen“ im Aufgabentyp „Vergleich“ mit einem Mittelwert von 0.062 (siehe Anhang). Nur sehr wenige Schülerinnen und Schüler erfüllten die Kriterien bei diesem Indikator. Die meisten Lernenden haben jedoch den Eiskwürfel gezeichnet und erfüllen dadurch den Indikator „Wurden beide Eiskwürfel dargestellt?“.

Tab. 38 Übereinstimmungen aller verglichenen Kriterien, aufgeschlüsselt nach Aufgabentyp und Kontext

Aufgabentyp	Kontext	Anzahl Indikatoren				
		total	mit $\kappa=1$	mit κ nicht definiert	mit $\kappa = 0$ bzw. $\kappa < 0$	mit $\kappa < 0.61$
Beobachtung	Brausetablette	12	5	0	0	1
	Eis schmelzen	11	2	1	0	1*
	Säuren und Basen	12	0	1	0	0
Vergleich	Brausetablette	20	0	2	3	1+ 2*
	Eis schmelzen	22	4	0	1	1
	Säuren und Basen	19	3	1	1	1
Experimentieren	Brausetablette	19	9	0	2	3
	Eis schmelzen	19	4	3	2	3
	Säuren und Basen	19	4	2	1	1

Über alle 153 Indikatoren (siehe auch Anhang) gesehen haben neun Indikatoren ein Cohens Kappa von Null ($\kappa = 0$) und weitere zwei Indikatoren ein negatives Cohens Kappa ($\kappa = -0.024$). Alle diese Indikatoren haben eine prozentuale Übereinstimmung von 92.5 % oder mehr. Ein Wert von Null des Cohens Kappa wird erreicht, wenn vorhandene Übereinstimmung mathematisch gesehen als rein zufällig angesehen werden kann. Ein negativer Wert weist auf eine Übereinstimmung hin, die kleiner ist als eine zufällige Übereinstimmung. Alle Indikatoren weisen jedoch ein Gwet AC1 (Gwet, 2008) von über 0.91 (siehe Anhang) auf und erreichen bei der Erfüllung des Indikators durch die Schülerinnen und Schüler einen Mittelwert über 0.9 (siehe Anhang). Dies zeigt die deutliche Asymmetrie im Antwortverhalten der Lernenden auf.

Die Tab. 38 zeigt eine Zusammenstellung der Übereinstimmungen aller verglichenen Kriterien, aufgeschlüsselt nach Aufgabentyp und Kontext. Die letzte Spalte weist die Anzahl Indikatoren mit einem ungenügenden Cohens Kappa ($0 < \kappa < 0.61$) aus (Cohen, 1960). Drei der Indikatoren haben zudem ein nicht genügendes Gwet AC1 ($\text{Gwet AC1} < 0.91$) (Gwet, 2008). Diese sind mit einem Stern (*) gekennzeichnet.

Tab. 39 Indikatoren mit ungenügender Objektivität

Aufgabentyp	Kontext	Facette	Indikator	Gwet AC1
Beobachtung	Eis schmelzen	fachliche Korrektheit	Sind Furchen/Einbuchtungen des Schmelzvorgangs mit dem Salz erkennbar?	0.828
Vergleich	Brausetablette	fachliche Korrektheit	Ist die Größe der Blasen korrekt dargestellt? i) Sind die Blasen beim Smartie kleiner als bei der Brausetablette?	0.794
		Unterschied erfasst und Vergleich dargestellt	Sind die relevanten Unterschiede bezüglich der Blasengröße deutlich dargestellt? i) Blasen beim Smartie deutlich kleiner?	0.794

9.2 % aller Indikatoren zeigen ein Cohens κ von $0 < \kappa < 0.61$. Bei einer hohen Asymmetrie im Antwortverhalten der Lernenden (mehrheitlich erfüllt oder mehrheitlich nicht erfüllt) haben Abweichungen in der Kodierung eine große Auswirkung auf das Cohens Kappa. Für drei dieser Indikatoren (knapp 2 %) fällt das Gwet AC1 unter 0.91 aus. Die anderen Indikatoren mit einem ungenügenden Cohens Kappa ($\kappa < 0.61$)

weisen ein Gwet AC1 von über 0.91 auf ($\text{Gwet AC1} > 0.91$). Die Interrater-Übereinstimmung lag bei 83.3 %.

Tab. 39 zeigt eine Zusammenstellung von Items mit ungenügenden oder nicht akzeptablen Werten bei Kappa und Gwet AC1. Alle weisen jedoch eine Interrater-Übereinstimmung von mehr als 83 % auf.

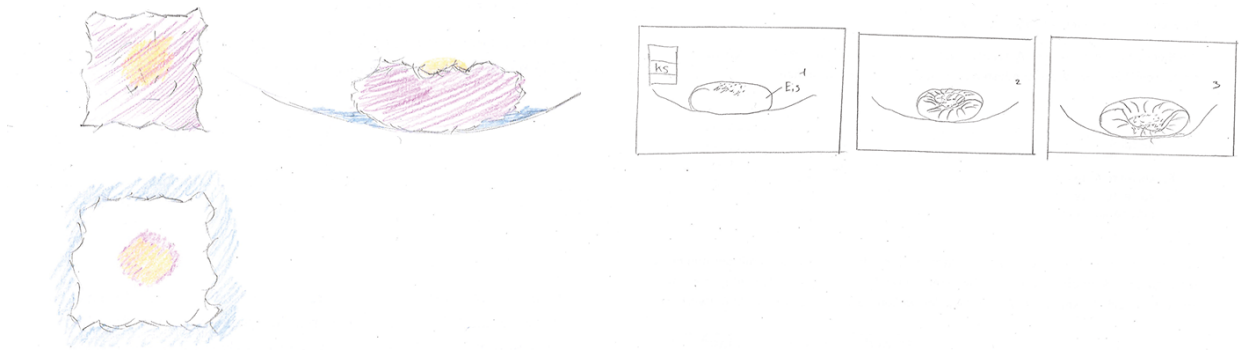


Abb. 6 Schülerskizzen des Kontexts „Eis schmelzen“ im Aufgabentyp „Beobachtung“ (EI_B)

Die drei sowohl mit dem Cohens κ als auch mit dem Gwet AC1 ungenügend ausfallenden Indikatoren werden von den Experten (Kodierer) als höchst inferent beurteilt. Beim Indikator „Sind Furchen/Einbuchtungen des Schmelzvorgangs mit dem Salz erkennbar?“ kann unklar sein, ob die Darstellung Furchen/Einbuchtungen zeigt oder ob es sich um die Darstellung von Salz handelt. Auch bei farbigen *wissenschaftlichen Skizzen* ist nicht klar erkennbar, ob die Farbänderungen die im Indikator gefragten Furchen/Einbuchtungen darstellen oder ob es sich um die Darstellung von Salz handelt (siehe Abb. 6 links). Der Indikator wurden dazu von den Lernenden oftmals schlecht dargestellt. Bei einem Mittelwert von nur 0.203 kann man von einem Bodeneffekt sprechen (Rost, 2004). In Abb. 6 rechts sind die Furchen/Einbuchtungen deutlich erkennbar dargestellt. Das Salz in der Mitte ist außerdem auch klar erkennbar dargestellt.

Die Abb. 7 zeigt vier Schülerskizzen des Kontexts „Brausetablette“ im Aufgabentyp „Vergleich“ (BR_V). Erkennbare Unterschiede im Löseverhalten von einem Smartie (immer links dargestellt) und einer Brausetablette (immer rechts dargestellt) sind die Ausbreitung der Blasen (bei der Brausetablette in die Breite, beim Smartie nur über dem Smartie) und die Blasengröße (beim Smartie sind die Blasen deutlich kleiner.) Die Blasengröße ist ein sehr inferenter Faktor, da die Blasen nicht immer durch kleine Kreise, wie in der Abb. 7 oben links, dargestellt werden. Ist die Blasengröße in der

Schülerskizze Abb. 7 oben rechts überhaupt erkennbar? Unterscheidet sich diese in der Darstellung mit der Brausetablette von derjenigen mit dem Smartie? Die beiden Schülerskizzen Abb. 7 unten zeigen zwei ganz unterschiedliche Darstellungen von Blasen. Die Übereinstimmung zwischen den Ratern liegt mit 83.3% zwar über den geforderten 80%, jedoch liegt der Wert des Gwet AC1 mit 0.7939 unter dem geforderten Wert von 0.91 (siehe Anhang). Die beiden Indikatoren sind stark inferent.



Abb. 7 Schülerskizzen des Kontextes „Brausetablette“ im Aufgabentyp „Vergleich“ (BR_V)

6.6.5 Auffällige Indikatoren

Die Indikatoren erfüllen unterschiedliche Aufgaben. Indikatoren mit einem Mittelwert von über 0.9 bspw. zeigen, dass fast alle Schülerinnen und Schüler diesen Indikator erfüllen bzw. nur gerade 10 % ihn nicht erfüllen. In einem solchen Fall spricht man von einem Deckeneffekt oder auch von einer großen Asymmetrie (Rost, 2004). Solche Indikatoren sind zwar für die Differenzierung von Kompetenzunterschieden nicht nützlich, aus inhaltlicher Perspektive jedoch notwendig, da sie beschreiben, was die Schülerinnen oder Schüler grundsätzlich gut beherrschen. Anders verhält es sich mit Indikatoren, die Mittelwerte zwischen 0.2 – 0.8 aufweisen. Sie geben nicht nur Aufschluss über die Leistungsdifferenzierung der einzelnen Niveaus im Progressionsmodell, sondern auch über die Kompetenzen der einzelnen Lernenden. Erfüllen nur wenige Schülerinnen und Schüler einen Indikator, dient der Indikator zur Beschreibung des höchsten Niveaus (Deckeneffekt).

Nachfolgend wird ein Überblick über die Erfüllung der Indikatoren und Beispiele für die unterschiedlichen Fälle gegeben sowie im Speziellen auf auffällige Indikatoren

eingegangen. Die Mittelwerte des Erfüllens der Indikatoren durch die Schülerinnen und Schüler liegen zwischen 0.062 und 1, wobei 24.8 % der Indikatoren einen Mittelwert von über 0.9 aufweisen. Nur 4.6 % haben einen Mittelwert von unter 0.2 (siehe Anhang).

Indikatoren, die in der Erfüllung einen sehr hohen Mittelwert erreichen, sind die Indikatoren der Facette „Idealisierung“ über alle Aufgabentypen und Kontexte hinweg. Fast alle Lernenden skizzieren die wesentlichen Materialien. Beim Aufgabentyp „Beobachtung“ im Kontext „Brausetablette“ sind das die Materialien Becher, Wasser und Brausetablette, die von über 95 % der Lernenden korrekt dargestellt wurden. Solche Indikatoren sind für die inhaltliche Beschreibung der Niveaus im Progressionsmodell hilfreich. Indikatoren der Facette „Vollständigkeit“ weisen über alle Aufgabentypen und alle Kontexte hinweg weder Decken- noch Bodeneffekte auf (Rost, 2004). Der Indikator „Wurde die Reaktion in mehreren Bildern dargestellt?“ (pro Indikator: eine genauere Beschreibung, z.B. Lösen der Brausetablette), der in allen Kontexten der beiden Aufgabentypen „Beobachtung“ und „Vergleich“ vorkommt, weist einen Mittelwert zwischen 0.417 und 0.762 auf. Indikatoren wie „Ist dargestellt, dass das Eis mit Salz schneller schmilzt?“ des Aufgabentyps „Beobachtung“ des Kontexts „Eis schmelzen“ und der Facette „Korrektheit“ zeigen Deckeneffekte. Nur 20.3 % der Schülerinnen und Schüler gelingt die Erfüllung des Indikators.

Tab. 40 Zusammenstellung Indikatoren mit niedrigen Mittelwerten des Aufgabentyps „Experimentieren“

Kontext	Facette	Indikator	Mittelwert
Brausetablette	Vollständigkeit	Wurde die Reaktion der Brausetablette nicht gezeichnet?	0.514
Säuren und Basen	Eindeutigkeit	Ist der Vorgang der Planung nur mit der Legende erkennbar und nachvollziehbar? ((Schwarz)tee, Zugabe von Seife und Zitronensaft zum Schwarztee)	0.492

Im Aufgabentyp „Experimentieren“ liegen die Mittelwerte zur Erfüllung eines Indikators im Schnitt deutlich höher als bei den anderen beiden Aufgabentypen. Ausnahmen bilden zwei Indikatoren (siehe Tab. 40), auf welche im Weiteren einzeln eingegangen wird.

Der Mittelwert des Erfüllens des Indikators von „Wurde die Reaktion der Brausetablette nicht gezeichnet?“ fällt im Kontext „Brausetablette“ markant niedriger aus als derjenige der äquivalenten Indikatoren bei den Kontexten „Eis schmelzen“ und „Säure und Base“. Grund dafür ist die rasche Reaktion der Brausetablette im Gegensatz zu den beiden anderen, etwas langsameren Reaktionen. In der Video-Vignette, welche die Lernenden sehen, wird die Reaktion der Brausetablette am Ende des Filmes kurz gezeigt, was bei den anderen beiden Kontexten nicht der Fall ist. Auch wenn vor jeder Bearbeitung darauf hingewiesen wird, dass die Reaktion nicht dargestellt werden soll, stellen nicht ganz die Hälfte aller Lernenden die Reaktion dennoch dar.

Der Mittelwert zur Erfüllung des Indikators der Facette „Eindeutigkeit“ fällt im Kontext „Säuren und Basen“ markant tiefer aus als derjenige der äquivalenten Indikatoren der beiden anderen Kontexte. Dies vor allem aus dem Grund, dass im Kontext „Säuren und Basen“ über alle drei Aufgabentypen die Legende für die Erfüllung dieses Indikators zwingend notwendig ist, während bei den anderen beiden Aufgabentypen darauf verzichtet wurde.

6.7 Validierung

Die Validierung gilt als ein zentrales Qualitätskriterium bei der Entwicklung von Leistungsmessinstrumenten. Messick (1995) unterscheidet sechs Validitätsaspekte, deren systematische Bewertung eine differenzierte Einschätzung des entwickelten Tests zulässt und darüber hinaus Hinweise für dessen Weiterentwicklung liefert (siehe auch (Miller & Linn, 2000)). „Diese sechs Validitätsaspekte erlauben eine systematische, alle Schritte einer Kompetenzmodellierung durchdringende Validitätsanalyse, die dabei hilft, die für jeden Schritt spezifischen Validitätsbedrohungen zu erkennen und zu bewältigen“ (Leuders & Sodian, 2013).

Der von Messick (1995) als inhaltliche Validität bezeichnete Aspekt bezieht sich auf die curricularen Gesichtspunkte und das Literaturstudium des modellierten Bereichs in Bezug auf die Kontexte (Content Aspect). Die inhaltliche Validität wird folgenden aufgeteilt in eine ökologische Validierung in Bezug auf die Kontexte wie auch in Bezug auf die Qualitätsmerkmale des Kompetenzstrukturmodells (Validierung Qualitätsmerkmale). Auf die Prüfung der kognitiven Validität, also die Passung der kognitiven Prozesse bei der Kompetenzerfassung zum postulierten theoretischen Kompetenzmodell (Substantive Aspect), wurde verzichtet. Die konsequentielle Validität, al-

so die Angemessenheit der Nutzung im pädagogischen oder bildungspolitischen Kontext (Consequential Aspect) wird im Ausblick (siehe Kapitel 8) angedacht.

6.7.1 Validierung der Qualitätsmerkmale des Kompetenzstrukturmodell

Die Validierung der Qualitätsmerkmale des Kompetenzstrukturmodells (siehe 3.5.2 und Tab. 8) wurde durch ein Expertenpanel (Crocker, 1997) aus Fachdidaktikern der Chemie sichergestellt. Im Januar 2020 wurden mit einem schriftlichen Fragenkatalog zwanzig Chemie-Fachdidaktiker und Chemie-Fachdidaktikerinnen aus den Ländern Schweiz und Deutschland zur Relevanz der Qualitätsmerkmale befragt. Die Items beinhalteten die Kriterien des *wissenschaftlichen Skizzierens*, wie „Vollständigkeit“ oder „zwei Phänomene oder Vorgänge vergleichen, indem sie nebeneinander gezeichnet werden“. Ergänzend bestand die Möglichkeit, weitere Kriterien aufzuführen. Erfasst wurde die Berufserfahrung als Lehrperson und die Berufserfahrung als Fachdidaktikerin bzw. Fachdidaktiker in Chemie. Die Befragten verfügten über eine Berufserfahrung zwischen einem bis vierzig Jahren in der Fachdidaktik und einer Berufserfahrung als Lehrperson von bis zu fünfzig Jahren (siehe Kapitel 7.2, F1 und Anhang).

6.7.2 Ökologische Validierung des Aufgabensets

Die curriculare Abdeckung der Kontexte im Aufgabenset wurde durch die Anlehnung an den Lehrplan 21 (D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz, 2016b) sichergestellt. Eine ökologische Validierung (Moosbrugger & Kelava, 2020) erfolgte durch eine Befragung von neunundzwanzig Lehrpersonen der Sekundarstufe I aus unterschiedlichen deutschsprachigen Regionen der Schweiz.

Die Befragung fand im Herbst 2019 statt. Sie wurde schriftlich an die Lehrpersonen versandt. Die Teilnahme erfolgte freiwillig. Der Fragebogen umfasste Fragen zur inhaltlichen Umsetzung von Kontexten im Unterricht, wie auch, ob diese Kontexte in dem im Unterricht verwendeten Lehrmittel vorkamen. Des Weiteren wurde ermittelt, ob die Aufgabentypen „Beobachtung“, „Vergleich“ und „Experimentieren“ im Unterricht der befragten Lehrpersonen und im verwendeten Lehrmittel vorkamen. Das Antwortformat beinhaltete vier Möglichkeiten von nie bis sehr oft. Erfragt wurde auch die Unterrichtserfahrung, die Jahrgangstufen und Niveaus, die die Lehrpersonen unterrichteten. Die Lehrpersonen verfügten über eine Berufserfahrung von einem bis zu vierzig Jahren und unterrichteten auf unterschiedlichen Niveaus. Sie arbeiteten mit teil

weise selbst zusammengestellten Unterrichtsmaterialien und mit unterschiedlichen Lehrmitteln der Sekundarstufe I (siehe Kapitel 7.3, F2 und Anhang).

6.7.3 Strukturelle Validierung

Unter der strukturellen Validität verstehen Leuders und Sodian (2013) die Passung von theoretischem Kompetenzmodell und gewähltem psychometrischem Messmodell (Structural Aspect). Die strukturelle Validität wird durch die Auswertung der Daten ermittelt, indem folgenden Fragen nachgegangen wird: Wird die Struktur der Kompetenzen durch die Daten abgebildet? Diskriminieren die Items genügend gut zwischen Personen mit geringen und hohen Fähigkeiten? Liegt eine genügend gute Varianz vor? Liegt eine Passung des Tests zur zu überprüfenden Fragestellung vor? Ebenso fallen Dimensionsanalysen unter die strukturelle Validierung (Leuders & Sodian, 2013). Die Testgüte wurde im Rahmen der strukturellen Validität ermittelt (siehe F3 und Kapitel 7.4.1).

6.7.4 Generalisierbarkeit

Die Generalisierbarkeit ist die Angemessenheit einer über die Aufgaben- und Personengruppe hinausgehenden Interpretation (Generalizability Aspect) (Leuders & Sodian, 2013). Können die erforschten Daten auf andere Anforderungssituationen übertragen werden? Haben Geschlecht, Jahrgangsstufe, Anzahl gesprochener Sprachen, Personenfähigkeiten wie auch der Schultyp einen Einfluss? Diesen Fragen wurde mit Hilfe von DIF-Analysen nachgegangen, mit dem Ziel, die Messinvarianz des Messinstruments zu überprüfen und dadurch Aussagen über die Generalisierbarkeit zu machen (siehe F3 und Kapitel 7.4.6).

6.7.5 Externe Validierung

In Abgrenzung zur internen Validität, die experimentelles Design, Randomisierung, Experimental- und Kontrollgruppe sowie die Kontrolle von Störfaktoren berücksichtigt, ist die externe Validität die Angemessenheit mit Blick auf konvergente, diskriminante und prädiktive Zusammenhänge mit anderen Konstrukten (External Aspect). Messmodelle beschreiben das Verhalten von Personen in bestimmten Situationen (Aufgaben, Items) in Abhängigkeit der Anforderungsmerkmale der Situationen (Items-Merkmale) und der Disposition der Personen (latente Personenmerkmale) (Leuders & Sodian, 2013). Rückschlüsse auf andere kognitive Fähigkeiten und Fertig-

keiten geben der KFT (Ermittlung der drei Intelligenz-Skalen), der LGVT (Lesefähigkeit) wie auch der NAW (Strategiewissen zum Experimentieren) (siehe Kapitel 7.5.1 und F4 und F6).

Beziehungen zum Fach- und Methodenwissen wurden durch Korrelationsberechnungen wie auch Strukturgleichungsmodelle ermittelt (siehe Kapitel 7.5.2 und F5).

6.8 Probabilistische Testtheorie

Die erhobenen Daten wurden nach der probabilistischen Testtheorie ausgewertet. Diese Theorie nimmt eine probabilistische Beziehung zwischen der Fähigkeitsausprägung einer Person (Personenparameter), der Itemschwierigkeit (Aufgabenparameter) und der Wahrscheinlichkeit, dass die Person die Aufgabe richtig löst, an. Angenommen wird, dass die Wahrscheinlichkeit, dass eine Person eine Aufgabe richtig löst, mit steigendem Personenparameter (Personenfähigkeit) und sinkendem Aufgabenparameter (Itemschwierigkeit) zunimmt (Rasch, 1993). Es gibt verschiedene probabilistische Modelle, mit denen sich auf der Grundlage der erhobenen Daten derartige Berechnungen anstellen lassen. Das sogenannte Rasch-Modell ist das einfachste Modell zur Schätzung von Itemschwierigkeiten und Personenfähigkeiten (wie gut ist eine Person in einer Kompetenz), da die Antworten der Lernenden bei einer Modellgeltung auf eine latente Variable zurückzuführen sind (Bühner, 2011; Rost, 2004). Notwendig zur Bestimmung dieser latenten Variablen ist die Bedingung, dass die manifesten Variablen, also die einzelnen Items, untereinander korrelieren und die latente Variable, also die Kompetenz, der Grund der Korrelation ist. Die Items dürfen nicht voneinander abhängig sein, d.h. u.a., dass sie außerdem unabhängig voneinander bearbeitbar sein müssen (Rost, 2004). Des Weiteren muss die Stichprobenunabhängigkeit (Itemparameter unabhängig von den Personen und Personenparameter unabhängig von den Items schätzen) gewährleistet sein. Im Folgenden wird auf die für die Arbeit relevanten Aspekte der Testtheorie eingegangen.

6.8.1 Das Rasch-Modell

Das Rasch-Modell basiert auf gewissen Annahmen, dazu gehört, dass die Wahrscheinlichkeit, dass eine Person eine Aufgabe löst in Abhängigkeit eines Personenparameters steht, jedoch unabhängig vom Aufgabenparameter ist (Rost, 2004). Das Modell ordnet dabei jeder Person einen Personenparameter und jeder Aufgabe einen Schwierigkeitsgrad (Aufgabenparameter) zu. Je höher der Personenparameter, desto

höher ist die Lösungswahrscheinlichkeit für ein Item mit einem festgelegten Aufgabenparameter. Als Modellannahme wird vorausgesetzt, dass das Ausmaß einer bestimmten Fähigkeit einer Person durch einen einzelnen Parameter charakterisiert wird und dass die Fähigkeit ein eindimensionales Merkmal ist. Der Schwierigkeitsgrad jeder Aufgabe wird durch einen einzigen Parameter charakterisiert und die Schwierigkeit ist ebenfalls ein eindimensionales Merkmal. Die Leistungen einer Person sind über alle Aufgaben hinweg unabhängig und hängen nur von den Itemschwierigkeiten, aber nicht von den zuvor gelösten Aufgaben ab. Die Antwortwahrscheinlichkeit wird im Funktionstyp zur Berechnung der Lösungswahrscheinlichkeit einer Person für eine Aufgabe in eine Logit-Funktion transformiert. Der Wertebereich reicht nun, anstelle von null bis eins, von minus unendlich bis plus unendlich, wobei allgemein Werte von -3 bis +3 Logits erwartet werden (Bühner, 2011). Es wird für die Logit-transformierte Lösungswahrscheinlichkeit (also die Funktion von Itemschwierigkeit und Personenfähigkeit) eine lineare Abhängigkeit des Personenparameters angenommen. Der Nullpunkt der Logitskala kann einer Lösewahrscheinlichkeit von 50 % entsprechen, dies hängt jedoch davon ab, wie die Itemparameter zu den Personenparametern stehen. Items und Personen mit einem negativen Logitwert haben niedrigere Schwierigkeiten bzw. niedrigere Personenfähigkeiten. Items und Personen mit einem positiven Logitwert hingegen haben eine hohe Itemschwierigkeit bzw. Personenfähigkeit (Bühner, 2011; Koller et al., 2012). Beide Parameter werden auf derselben Skala abgebildet (Rost, 2004). Die Wahrscheinlichkeit, dass eine Person v die Aufgabe i korrekt löst wird gezeigt durch die Formel 1:

Formel 1 Rasch-Modell-Gleichung (Koller et al., 2012)

$$P(X_{vi} = x_{vi} | \theta_v, \beta_i) = \frac{\exp[x_{vi}(\theta_v - \beta_i)]}{1 + \exp(\theta_v - \beta_i)} = p_{vi}$$

Die Exponentialfunktion, mit der sich jede Aufgabe darstellen lässt, bestimmt den Verlauf der Itemcharakteristikkurve (ICC; engl.: item characteristic curve) und bildet die Lösewahrscheinlichkeit für ein Item in Abhängigkeit von Personenfähigkeit und Itemschwierigkeit ab. Auf der y-Achse wird die Wahrscheinlichkeit der Itemerreichbarkeit dargestellt. Sie reicht von 0 bis 1. Die x-Achse repräsentiert die latente Variable, also die mögliche Personenfähigkeit (Rost, 2004).

Die Trennschärfe ist ein Qualitätskriterium der einzelnen Items. Sie gibt an, wie stark die mit dem jeweiligen Item erzielte Differenzierung zwischen den Testpersonen mit

der Differenzierung durch den Gesamttest übereinstimmt. Eine gute Trennschärfe eines Items lässt die Voraussage zu, dass Personen mit einem hohen Testergebnis bei einem gegebenen Item einen hohen Wert erzielen (Koller et al., 2012). Die Parallelität der ICCs, die als Merkmal der Rasch-Modelle gilt, basiert auf der Annahme, dass alle Items dieselbe Trennschärfe aufweisen (Rost, 2004). Die ICC-Kurve zeigt, dass bei konstanter Itemschwierigkeit die Lösungswahrscheinlichkeit eines Items mit steigender Personenfähigkeit zunimmt (Koller et al., 2012).

Die zur Berechnung eingesetzte Software ConQuest[®] verwendet für die Schätzung der Itemparameter die sogenannte marginale Maximum-Likelihood-Methode (Wu et al., 2007). Für die Schätzung der Itemschwierigkeit und der Personenfähigkeit wird von einer Normalverteilung der Population ausgegangen. Die Itemparameter werden so geschätzt, dass das Produkt der Wahrscheinlichkeiten der einzelnen Parameter (Likelihood L) möglichst groß wird (Bühner, 2011; Koller et al., 2012).

6.8.2 Das „Partial Credit Model“

Das in Kapitel 6.8.1 beschriebene Rasch-Modell dient als Grundlage für das Partial Credit Model. Das Partial Credit Model geht wie das Rasch-Modell davon aus, dass über die gegebenen Itemantworten auf eine gemeinsame, latente Variable geschlossen werden kann. Das Partial Credit Model berücksichtigt auch nichtdichotome Bewertungsformate, also Formate mit mehr als zwei Antwortmöglichkeiten. Das Partial Credit Model ist bei Variablen mit geordneten Antwortkategorien und damit auf ordinale Daten anwendbar. Die einzelnen Kategorien werden als Leistungsniveaus bezeichnet. Von diesen klar abzugrenzen sind die (Lösungs)schwellen (Thresholds), die jeweils zwischen zwei benachbarten Kategorien liegen. Durch die Bewältigung der ersten Stufe gelangt man auf das erste Leistungslevel, und durch Meistern der zweiten Stufe gelangt man auf das zweite Level, usw. Der Begriff der Schwelle impliziert, dass der Übergang von einer Kategorie zur anderen dem Kontinuum an diesem Punkt auf stattfindet, d. h. die Wahrscheinlichkeit, in der folgenden Kategorie zu antworten, von diesem Punkt an grösser wird als die Wahrscheinlichkeit, in der vorangegangenen Kategorie zu antworten (Rost, 2004). Auf der jeweiligen Schwelle, d.h. auf der Skala der latenten Variablen, haben die beiden betrachteten Kategorien exakt dieselbe Wahrscheinlichkeit. Damit besteht an diesem Punkt also jeweils eine 50 %-ige Wahrscheinlichkeit, auf das niedrigere Level $x-1$ oder auf das höhere Level x zweier benachbarter Niveaus zu gelangen. Die Schwellen liegen auf dem latenten Kontinuum

an den Schnittpunkten der Kategorienfunktionen (Rost, 2004). Je weiter rechts diese liegen, desto schwieriger ist es, sie zu überschreiten.

Von geordneten Antwortkategorien wird gesprochen, wenn die Schnittpunkte entlang des latenten Kontinuums angeordnet sind. Geordnet bedeutet dabei, dass die Antwortkategorien der Ratingskala gleich geordnet sind wie die entsprechenden Schwellenwerte. Die Schwellenparameter definieren Abschnitte auf dem Kontinuum, in denen die Testpersonen mit den entsprechenden Parametern für ein Item am wahrscheinlichsten antworten (Dorsch et al., 2013). Der Schwellenparameter zeigt an, wann einer Person wahrscheinlich ein höherer Code zugewiesen wird. Die Itemschwierigkeit wird als Mittelwert aller Schwellenwerte definiert (Bühner, 2011). Jedes Item verfügt über seinen eigenen Schwellenwert. Dies kann dazu führen, dass die verwendete Antwortskala bei Items ungeordnete Kategorien aufweist. Grund dafür ist, dass der Schwellenwert diejenige Fähigkeit angibt, ab der es wahrscheinlich ist, dass jemand von zwei benachbarten Codes den höheren erreicht. Falls Schwierigkeiten von benachbarten Codes sehr nahe beisammen sind, kann es vorkommen, dass ein Code für keine Kompetenzausprägung der wahrscheinlichste Fall ist (Rost, 2004). Die Kategorienfunktion dieses Codes wird überall durch die Kategorienfunktionen der benachbarten Codes übertroffen. Dies führt zu einer Verletzung der angenommenen Ordnung (Gut, 2012). Ein Beispiel dafür zeigt Rost (2004, S. 200). Um dieses Problem zu lösen, wird mit „Generalized Thresholds“ (verallgemeinerten Schwellenwerten) gearbeitet. Ein Generalized Threshold ist der Schwellenwert für diejenige Fähigkeit, ab der es wahrscheinlich ist, dass jemand einen höheren Code erreicht. Gemäß Definition sind die „Generalized Thresholds“ immer gleich geordnet. Während Rost (2004) die Ordnung der Schwellenparameter als Kriterium zur Überprüfung der Qualität der Item-Antwort-Ordinalskala sieht, weisen Masters und Wright (1997) darauf hin, dass neben der Abbildung der Antwortkategorien als Kompetenzbereiche, in denen ein Code der wahrscheinlichste ist, noch andere Abbildungsmöglichkeiten bestehen. Eine Abbildung, bei der die Interpretation der Schwellenparameter zudem invariant gegenüber einer Vertauschung der normalen Ordnung der Schwellenparameter ist, teilt sich das Kontinuum der Personenparameter in Bereiche ein, in denen die Wahrscheinlichkeit für einen höheren Code grösser ist als für einen tieferen Code (Masters & Wright, 1997).

6.9 Regressionsanalyse

Die Beziehung zwischen den Personenfähigkeiten im *wissenschaftlichen Skizzieren* und anderen Personenfähigkeiten wie der Kognition (KFT), der Lesefähigkeit (LGVT) oder dem Strategiewissen zum Experimentieren (NAW) wird durch eine Regressionsanalyse untersucht. Eine Korrelation zeigt dabei einen Zusammenhang an zwischen dem *wissenschaftlichen Skizzieren* und anderen Konstrukten (Bühner & Ziegler, 2009).

Für die Arbeit wurden folgende Voraussetzungen der Berechnung von linearen Regressionsmodellen getestet:

- a) Linearität, insbesondere auch keine Ausreisser
- b) keine Autokorrelation der Residuen
- c) Homoskedastizität (Gleichheit der Varianzen der Residuen)
- d) normalverteilte Residuen
- e) keine Multikollinearität, d.h. keine Korrelationen der Prädiktoren mit externen Variablen (Bortz & Schuster, 2010; Field, 2018).

Als Nächstes werden die verschiedenen Voraussetzungen, deren Testung und die Resultate kurz aufgeführt. Die Überprüfung der Autokorrelation erfolgte mit der Durbin-Watson-Statistik, die Werte von 0 bis 4 annehmen kann. Ein Wert von 2 besagt, dass keine Autokorrelation zwischen den Residuen besteht. Für Werte nahe bei 2 kann davon ausgegangen werden, dass ebenfalls keine Autokorrelation vorliegt (Bühner & Ziegler, 2009). Hier beträgt der Wert der Durbin-Watson-Statistik 1.94, das heißt es liegt keine Autokorrelation vor.

Die Residuen, Vorhersagefehler, d.h. die Differenz zwischen beobachteten Werten und den durch die Regressionsgeraden vorhergesagten Werten, sind normalverteilt.

Homoskedastizität, d.h. die Gleichheit der Varianzen der Residuen, liegt vor, wenn die Residuen für alle Variablen gleich sind. Die Prüfung erfolgte über Streudiagramme mit den z-standardisierten vorhergesagten Werten als x-Achse und den standardisierten Residuen als y-Achse. Bei einer Homoskedastizität streuen die vorhergesagten Werte gleichmäßig um die Regressionsgerade (Bühner & Ziegler, 2009). Auch die Normalverteilung der Residuen, eine weitere Voraussetzung, die erfüllt sein muss (Field, 2018) liegt vor. Sie wurde mittels Histogrammen der standardisierten Residuen zur visuellen Überprüfung begutachtet (Bühner & Ziegler, 2009).

Die Überprüfung der Multikollinearität erfolgte über den „Variance Inflation Faktor“ (VIF) und die Toleranz. Ein Wert unter 0.1 für die Toleranz sowie ein Wert über 10 für den VIF ist starkes Zeichen für die Multikollinearität (Field, 2018). Mit Werten für die Toleranz von 0.741 und einem VIF von 1.35 kann eine Multikollinearität wie auch die Unklarheit, welcher Prädiktor für die Varianzaufklärung relevant ist, ausgeschlossen werden. Beides hätte Auswirkungen auf die Standardfehler der β -Gewichte. Hohe β -Gewichte würden nicht signifikant werden, tiefe dagegen schon. Multikollinearität tritt auf, wenn zwei Prädiktoren miteinander stark korrelieren (Bühner & Ziegler, 2009).

Zur Berechnung der externen Validität der Kompetenzmessung wurde eine Regressionsanalyse für die Kompetenz des *wissenschaftlichen Skizzierens* mit den Begleitstudien (NAW, KFT und LGVT, siehe Kapitel 6.3.5 und 7.5) ermittelt. Die Analyse der linearen Regression erfolgte nach Transformation der Daten aus dem Statistikprogramm ConQuest® in SPSS® (Wu et al., 2007). Die Regression wird als lineares Modell aus den Messdaten ermittelt (Field, 2018). Beim unstandardisierten Regressionsgewicht b oder B handelt es sich um die Steigung der Regressionsgeraden. Diese gibt an, um wieviel sich die abhängige Variable ändert, wenn sich die unabhängige Variable um eine Einheit ändert (Field, 2018). Das standardisierte Regressionsgewicht β wird vom unstandardisierten abgeleitet, indem die Regressionsgerade für z -standardisierte Werte dargestellt wird und der y -Achsenabschnitt dadurch wegfällt, da dieser Teil der Wirkung der z -Transformation unwichtig für die Umrechnung von b nach β ist. Der β -Wert gibt wie das unstandardisierten Regressionsgewicht (b oder B) an, dass sich der Wert des Kriteriums um β -Standardabweichung ändert, wenn sich der Wert des Prädiktors um eine Einheit ändert (Bühner & Ziegler, 2009). Ein Regressionsgewicht von $\beta = 0$ bedeutet keinen Zusammenhang, hohe β -Werte weisen auf große Zusammenhänge zwischen den untersuchten Merkmalen hin.

Cohen (1988) legt für die Interpretation der Effektstärke des Zusammenhangs folgende Werte fest: Beta (β) > 0.4 : starker Zusammenhang, $0.2 - 0.4$: moderater Zusammenhang und $0.0 - 0.2$: kein oder schwacher Zusammenhang. Nach Peterson und Brown (2005) lässt sich das standardisierte Beta (β) mit dem R gleichsetzen; daher kann die Varianzaufklärung für die einzelnen Prädiktoren geklärt werden.

Der Determinationskoeffizient R^2 gibt Aufschluss über den Anteil der Varianz, die durch das Modell an der Gesamtvarianz aufgeklärt werden kann (Bühner & Ziegler,

2009). Bei einer multiplen Regression können die Korrelationen unter den Prädiktoren bei der Berechnung der Varianz zur Aufklärung der Varianz der einzelnen Prädiktoren mit einbezogen werden (Moosbrugger & Kelava, 2012). Der korrigierte Determinationskoeffizient R^2_{kor} gibt an, wie gut die Modellfunktion die Daten erklärt. Der Korrelationskoeffizient R zeigt die Stärke des linearen Zusammenhangs zwischen zwei quantitativen Größen. Für die Interpretation des korrigierten Determinationskoeffizienten R^2_{kor} wird ein Wert von $R^2_{\text{kor}} < 0.02$ als kleiner, ein Wert $R^2_{\text{kor}} < 0.13$ als mittlerer und ein Wert $R^2_{\text{kor}} < 0.26$ als grosser Effekt klassifiziert (Cohen, 1988).

Die Effektstärke f der Prädiktoren lässt sich nach Cohen (1988) aus R^2 mit der Formel 2 ermitteln:

Formel 2 Effektstärke der Regressionsanalyse

$$f = \frac{R^2}{1-R^2}$$

Für die Interpretation der Effektstärke wird ein Wert $f < 0.02$ als kleiner, ein Wert $f < 0.15$ als mittlerer und ein Wert $f < 0.35$ als großer Effekt klassifiziert (Cohen, 1988).

6.10 Korrelationsanalyse

Bei der Erhebung der Daten wurden den Lernenden jeweils sechs Items vorgelegt. Jedes Item bestand aus einer Fragestellung zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* mit dem Antwortformat „*wissenschaftliche Skizze*“ und einem Set von Fragen zum Fach- und Methodenwissen in Form eines Multiple-Select Tests (siehe Kapitel 6.3.4), der auf das jeweilige Item abgestimmt wurde. Für den Aufgabentyp „Beobachtung“ waren es drei Aufgaben mit jeweils einer korrekten Antwort und zwei Distraktoren. Der Kontext „Säuren und Basen“ verfügte über zwei korrekte Antworten und einen Distraktor. Beim Aufgabentyp „Vergleich“ gab es vier Aufgaben mit einer korrekten Antwort und zwei Distraktoren für den Kontext „Brausetablette“. Für die Kontexte „Eis schmelzen“ und „Säure und Basen“ hingegen gab es zwei korrekte Antworten und einen Distraktor. Beim Aufgabentyp „Experimentieren“ wurden zwei Aufgaben gestellt mit je sechs Antwortmöglichkeiten, wobei davon jeweils zwei korrekt waren und vier Distraktoren darstellten.

Korrelationen dienen der Analyse von Zusammenhängen zwischen Variablen. Rangkorrelationen, wie der hier angewendete Kendall-Tau-Test, benötigen nicht die Annahme, dass die Beziehung zwischen den Variablen linear ist. Dies macht den Test

robust gegenüber Ausreißern (Field, 2018). Beim Kendall-Tau-Test wird von ordinalskalierten Daten ausgegangen. Der Test misst, wie oft die Rangfolge des Merkmals A nicht mit der Rangfolge der Merkmale B übereinstimmt, und dividiert diese durch die prinzipiell möglichen Rangfolgen, wodurch das Intervall des Kendall-Tau-Tests zwischen -1 und 1 liegt. Ein Absolutwert von 1 besagt eine identische Rangfolge und somit einen vollkommenen Zusammenhang, und ein positiver Wert zeigt einen positiven Zusammenhang auf. Bei einer hohen Ausprägung des einen Merkmals kann auf eine hohe Ausprägung des anderen Merkmals geschlossen werden.

Ein negativer Wert für die Korrelation weist auf einen negativen Zusammenhang hin (Bortz & Döring, 2007). Dies bedeutet, dass bei einer hohen Ausprägung des einen Merkmals eine niedrige Ausprägung des anderen Merkmals vorliegt. Bei einem Korrelationswert von Null liegt kein Zusammenhang zwischen den Merkmalen vor. Eine Einstufung der Korrelationskoeffizienten bietet Zöfel (2003) (siehe Tab. 41).

Tab. 41 Einstufung der Korrelationskoeffizienten (Zöfel, 2003)

Korrelationskoeffizienten	Einstufung
$ r \leq 0.2$	sehr geringe Korrelation
$0.2 \leq r \leq 0.5$	geringe Korrelation
$0.5 \leq r \leq 0.7$	mittlere Korrelation
$0.7 \leq r \leq 0.9$	hohe Korrelation
$0.9 \leq r \leq 1.0$	sehr hohe Korrelation

6.11 Strukturgleichungsmodell

Strukturgleichungsmodelle dienen dazu, theoretisch hergeleitete Kausalzusammenhänge zwischen latenten Variablen mit empirischen Daten zu überprüfen (Reinecke & Pöge, 2010). Multiple Regressionsanalysen stellen die einfachste Modellgleichung dar. Im Unterschied zu den Regressionsanalysen können gleichzeitig mehrere Hypothesen bezüglich der Beziehungen zwischen unabhängigen und abhängigen manifesten Variablen in einer Analyse überprüft werden. Die Verknüpfung mehrerer Regressionsmodelle miteinander wird bei der Modellierung als Pfadanalyse bezeichnet (Reinecke & Pöge, 2010). Pfaddiagramme dienen der graphischen Darstellung von Strukturgleichungsmodellen (Geiser, 2011). Mittels konfirmatorischer Faktorenanalyse wird im allgemeinen Strukturgleichungsmodell zwischen manifesten und laten-

ten Variablen differenziert. Dadurch ist die Kontrolle von Messfehlern möglich (Reinecke & Pöge, 2010). Geschätzt werden die Parameter durch die Minimierung von Diskrepanzfunktionen. Oftmals handelt es sich dabei um eine Maximum-Likelihood-Diskrepanzfunktion (Reinecke & Pöge, 2010).

Die Berechnung des Strukturgleichungsmodells erfolgte nach Transformation der Daten aus dem Programm SPSS® in MPlus®. Die Modellpassung kann anhand verschiedener Kriterien zur Beurteilung der Güte des Strukturgleichungsmodells vorgenommen werden. Der sogenannte Comparative-Fit-Index (CFI) wie auch der Tucker-Lewis-Index (TLI) sollten beide bei 0.97 oder noch näher bei 1 liegen. Diese beiden Parameter vergleichen das Zielmodell mit einem Unabhängigkeitsmodell (Geiser, 2011). Die informationstechnischen Masse (AIC, BIC) dienen zum Vergleich verschiedener Modelle auf deskriptiver Ebene. Als Entscheidungskriterium gilt, dass das Modell mit dem kleinsten Wert ausgewählt wird (Geiser, 2011).

7 Resultate und Interpretationen

Ziel des Dissertationsprojekts ist die Entwicklung und Validierung eines Messinstruments, das aus einem literaturbasierten, a-priori Kompetenzstrukturmodell und auf einer anschließenden externen Validierung des Messinstruments besteht. Die erfassten Studiendaten wurden aufbereitet (siehe Kapitel 7.1) und ausgewertet.

Nachfolgend werden die Resultate der Validierung der Qualitätsmerkmale des Kompetenzstrukturmodells und deren Interpretation (siehe Kapitel 7.2, F1), die ökologische Validierung des Testinstruments (siehe Kapitel 7.3, F2), die Validierungsaspekte der Testgüte des Messinstruments sowie die Verallgemeinerung (siehe Kapitel 7.4, F3) vorgestellt. Zusammenhänge zu externen Faktoren, die außerdem für die externe Validierung überprüft wurden, werden ebenfalls erläutert (siehe Kapitel 7.5, F4-F6).

7.1 Datenaufbereitung

Die *wissenschaftlichen Skizzen* der Lernenden wurden mit Hilfe des Kodiermanuals (siehe Anhang) in das im Progressionsmodell (siehe Kapitel 6.2) festgelegte Niveau des *wissenschaftlichen Skizzierens* eingeteilt, indem jede Skizze mit Hilfe von Indikatoren dichotom (null oder eins) kodiert wurde. Bei fehlenden Skizzen wurden alle Indikatoren mit „fehlend“ (999) kodiert. Mit der Software Excel® wurde aus den Daten der einzelnen Indikatoren das Niveau der Kompetenz des *wissenschaftlichen Skizzierens* (Niveau 0 bis 3) ermittelt. Das jeweilige Niveau wurde durch die im Kodiermanual festgehaltenen Schwellenwerte festgelegt (siehe Kapitel 6.5.2).

Die Daten des begleitenden Multiple-Select Tests (MS) (siehe Kapitel 6.3.4) wurden manuell ausgewertet. Pro vollständig korrekt beantwortete Frage wurde ein Punkt vergeben (alle Kreuze einer Frage mussten richtig gesetzt sein). Die Leistung der Schülerinnen und Schüler im Lesegeschwindigkeits- und Verständnistest (LGVT) (siehe Kapitel 6.3.5) wurde in Excel® erfasst und nach Schneider, Schlagmüller und Ennemoser (2007) ausgewertet. Die Items mit geschlossenem Antwortformat (NAW und KFT) (siehe Kapitel 5.3.5) wurden mit der Software Teleform®(Cardiff, 2008) eingelesen und anschliessend in SPSS®-Dateien gespeichert. Die Auswertung an der Anzahl korrekt gelöster Aufgaben wurde im Excel® vorgenommen.

Der Lesegeschwindigkeits- und Verständnistest (LGVT) mit den Subskalen (Leseverständnis LV, Lesegenauigkeit LGN und Lesegeschwindigkeit LGS) besitzt eine sehr

gut Testgüte, die durch Paralleltests und Retests ermittelt wurde (Schneider et al., 2007); diese wurde durch ein Cronbach-Alpha (α) von 0.728 für den Lesegeschwindigkeits- und Verständnistest mit einer zufriedenstellenden Übereinstimmung für die interne Konsistenz der Skala bestätigt. Die drei Skalen ergaben manifeste T-Werte, die ins Hintergrundmodell eingeflossen.

Das Cronbach-Alpha (α) für die Subskala zum nonverbal-figuralen Denken (N3) des KFT (Heller & Perleth, 2000) zeigt mit 0.796 eine zufriedenstellende Übereinstimmung für die interne Konsistenz der Skala (Bortz & Döring, 2007; Schmitt, 1996).

Für die Analyse der Items des NAW-Tests (Koenen, 2014; Mannel, 2011) wurde eine Rasch-Analyse durchgeführt. Die Berechnung der Skala erfolgte eindimensional. Zur Analyse der Aufgabenparameter wurden die gewichteten (Infit) und die ungewichteten (Outfit) Mean-Square-Fitstatistiken (MNSQ) ermittelt, und die Überprüfung erfolgte anhand eines Tests (T-Test) mit Hilfe z-standardisierter Werte. Werte zwischen 0.8 und 1.2 gelten für die Infit- und Outfit-Werte als gut (Wu et al., 2007). T-Werte sollten zwischen - 2.0 und 2.0 liegen (Wu et al., 2007). Als Signifikanztest wurde ein χ^2 -Test eingesetzt (Wu et al., 2007).

Tab. 42 Item-Kennwerte der neun Items des NAWs der Rasch-Analyse

N	375
Infit-Werte	0.90 < infit < 1.11
Outfit-Werte	0.91 < outfit < 1.04
T-Wert	-2.20 < T < 0.8
EAP/PV Reliabilität	0.713
Item Separation-Reliabilität	0.981
Variance	1.226
Geschätzte Parameter	10
Discrimination	0.47-0.58

Die Personen-Separation-Reliabilität liegt bei 0.713, die Item-Separation-Reliabilität ist mit 0.981 sehr zufriedenstellend (Wu et al., 2007). Die Varianz liegt mit 1.226 in einem hohen Bereich. (Wu et al., 2007). Die Rasch-Analyse zeigt für die neun eingesetzten Items eine zufriedenstellende Modellpassung (siehe Tab. 42) (Wu et al., 2007).

Die Datensätze der Begleitstudien NAW, KFT und LGVT liegen alle normalverteilt vor.

7.2 Validierung der Qualitätsmerkmale des Kompetenzstrukturmodells (F1)

Zur Beantwortung der F1 „Sind die Qualitätsmerkmale des entwickelten Kompetenzstrukturmodells inhaltlich valide?“ (siehe auch Kapitel 3.5.2 und 6.7.1) wurden die Inhalte des Kompetenzstrukturmodells und somit die Grundanforderungen an die Kompetenz des wissenschaftlichen Skizzierens validiert. Die Antworten der befragten Chemie-Fachdidaktikerinnen und -Fachdidaktiker sind in der Abb. 8 dargestellt. Sehr hohe Nennungen erhalten die Gruppe „Eindeutigkeit“ (wahrnehmbare relevante Qualitätsmerkmale dargestellt), „Fokus/Relevanz“ und „fachliche Korrektheit“. Dagegen werden „Detailtreue“ wie auch „3D-Darstellung“ als unwichtig erachtet.

Bei „Ergänzung durch Text“ herrscht mit etwas mehr Nennungen „eher ja“ als „eher nein“ keine Einigkeit unter den Fachdidaktikerinnen und Fachdidaktikern. Die Gruppe „Aufgabentyp“ zeigt im Schnitt weniger positive Nennungen, wahrscheinlich, da sie spezifisch die Aufgabentypen aufgreift.

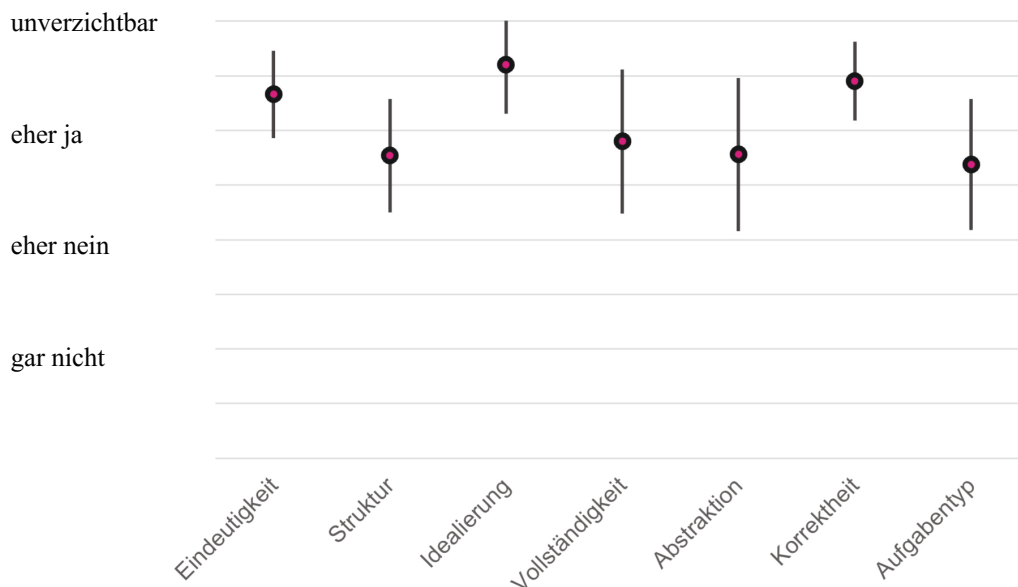


Abb. 8 Validierung der Qualitätsmerkmale wie auch des Aufgabentyps: „unverzichtbar“ = 4, „eher ja“ = 3, „eher nein“ = 2, „gar nicht“ = 1

Tab. 43 Validierung der Qualitätsmerkmale und des Aufgabentyps

Eindeutigkeit	deutlich gezeichnete Skizze	3.40
	wahrnehmbare relevante Qualitätsmerkmale dargestellt (Farbveränderungen, Formveränderungen, Aggregatzustandsänderungen etc.)	3.85
	Beschriftung/Legende verwendet	3.55
	Ergänzung durch Text	2.65
Struktur	angemessene Größe	2.75
	relevante Proportionen	2.85
	Perspektive wechseln	2.58
	Zoom auf Details	2.42
	Symbole (z. B. Pfeile, um Vorgänge darzustellen)	3.50
Idealisierung	Irrelevantes weggelassen	3.50
	Fokus/Relevanz (Unwichtiges weggelassen, Wichtiges im Zentrum)	3.70
Vollständigkeit	Vollständigkeit	2.85
	Alle Vorgänge/Abfolgen gezeichnet	2.90
	Verknüpfungen zwischen zwei Skizzen hergestellt	2.95
Abstraktion	Abstraktion	3.22
	2D gezeichnet	2.65
fachliche Korrektheit	Fachlichkeit (inhaltlich korrekt)	3.70
	Unterscheiden Beobachtung und Erklärung	3.39
Aufgabentyp	Vergleich dargestellt	3.06
	Unterschied erfasst	3.19
	zwei Phänomene oder Vorgänge wurden, verglichen diese nebeneinander gezeichnet	3.16
	Phänomen wird durch Planung erzeugt	2.36
	Variablenkontrolle dargestellt	2.88
Idealisierung	Detailtreue	2.40
Abstraktion	3D-erkennlich	2.00

Tab. 43 zeigt Mittelwerte der einzelnen Nennungen der Validierung Qualitätsmerkmale, unterteilt in Facetten und den Aufgabentyp. Die Nennung „unverzichtbar“ wurde als 4, die Nennung „eher ja“ als 3, die Nennung „eher nein“ als 2 und die Nennung „gar nicht“ als 1 bewertet. Ein Wert von über 3.0 lässt auf darauf schließen, dass dieses Kriterium von einer Mehrzahl der Fachdidaktiker und Fachdidaktikerinnen als relevant erachtet wird.

Neben der Bewertung der vorgegebenen Qualitätsmerkmale bestand für die Fachdidaktikerinnen und Fachdidaktiker die Möglichkeit, weitere Kriterien und Qualitätsmerkmale, die sie für die Kompetenz des *wissenschaftlichen Skizzierens* als relevant betrachteten frei zu äußern. Einzelne Fachdidaktikerinnen und Fachdidaktikern erachteten die in der Tab. 44 aufgeführten Kriterien und Qualitätsmerkmale neben den in der Tab. 43 aufgeführten Qualitätsmerkale als relevant.

Tab. 44 Kriterien und Qualitätsmerkmale Nennungen einzelner Fachdidaktik Expertinnen und -experten

Kategorie	Nennungen
Struktur	Klarheit Verständlichkeit Strukturiertheit gleiche oder ähnliche Beobachtungen mit gleichen Symbolen dargestellt
Vereinfachung	komplexe Zusammenhänge vereinfachen
Modellarbeit	Modellcharakter der Skizze soll sichtbar sein Unterscheidung sichtbare Welt / submikroskopische Denkmodellwelt kein Zoom von der sichtbaren Welt in die Denkmodellwelt, da sonst unklar bleibt, wann Denkmodellwelt startet bzw. unklar ist, dass die Eigenschaften sich ebenfalls vom sichtbaren Stoff zum Atom/Molekül/Teilchen ändern
Prozess/Ablauf	Entwicklung und Prozesse darstellen Anfangs- und Endzustände immer ganz klar dargestellt
Nutzen von Skizzen	Verzicht auf eine Skizze, wenn sie nicht dem Verständnis der Sache/des Phänomens dient

Die Nennungen der Kategorie „Struktur“ werden nicht explizit im Kompetenzstrukturmodell aufgegriffen, sie sind jedoch die Grundlage für die Facette „Eindeutigkeit“ wie auch für die Verständlichkeit der Skizze im Allgemeinen. Die Nennung „gleiche oder ähnliche Beobachtungen werden mit gleichen Symbolen dargestellt“ weist auf die Verwendung einer einheitlichen „Bild/Zeichensprache“ hin, die analog in der schriftlichen Beschreibung von Vorgängen und Beobachtungen als Verwendung einer einheitlichen Fachsprache aufgefasst wird (Gut & Mayer, 2018).

Die Vereinfachung findet sich teilweise in den Facetten „Idealisierung“ und „Abstraktion“ (siehe Tab. 7) wieder, auch in der Facette „Vollständigkeit“, wenn es um die Darstellung von Abläufen geht. Die Nennungen der Kategorie „Prozess/Ablauf“ sind in der Facette „Vollständigkeit“ in den Indikatoren aufgegriffen, die die Verwendung

von mehreren Bildern (Bildabfolgen) generell, aber auch explizit für die Darstellung des Phänomens kodieren.

Die Modellarbeit wurde explizit aus dieser Arbeit ausgeklammert, da dies eine weitere Kompetenz, die „Kompetenz der Modellarbeit“, aufgreifen würde und die Trennung von der Kompetenz des *wissenschaftlichen Skizzierens* in den Skizzen dadurch erschwert würde. Auch auf den Nutzen der „Verwendung von Skizzen“ wurde in dieser Arbeit nicht eingegangen, weil vor allem auf die Nutzung der Skizze fokussiert wurde.

7.3 Ökologische Validierung (F2)

Zur Beantwortung der F2 „Sind die für die Testkonstruktion verwendeten Kontexte und Aufgaben ökologisch valide?“ (siehe auch Kapitel 6.7.2) wurden die Relevanz wie die auch die Anwendungen der Kontexte und der Aufgabentypen im Chemieunterricht der Sekundarstufe I validiert. Wie häufig die Kontexte und Aufgabentypen im Chemieunterricht der Sekundarschule I Anwendung finden, konnten die befragten Lehrpersonen mit einer Abstufung von „nie“, „selten“, „ab und zu“ und „regelmäßig“ nennen.

Bei über der Hälfte der Lehrpersonen fanden die folgenden Kontexte im Unterricht regelmäßig Anwendung: „Säuren und Basen“ wie auch explizit „Indikatoren“, „Phasenübergänge“, insbesondere „Schmelzen von Eis“, „einfache chemische Reaktionen“, „Kohlenstoffdioxid“ und „Chromatographie“. Die folgenden expliziten Anwendungen der Themen wie „Farbstoff-Chromatographie“ (Teil der Chromatographie), „Lösen von Brausetablette“ (Teil der chemischen Reaktion) wie auch „die Entstehung von Kohlenstoffdioxid aus Backpulver und Essig“ (Teil des Kohlendioxids) fanden nur bei einem Viertel der befragten Lehrpersonen regelmäßige Anwendung im Unterricht. Die Methodenkompetenz fasst die drei Aufgabentypen zusammen.

Der Aufgabentyp „Beobachtung“ fand bei mehr als 90 % der Befragten regelmäßig im Unterricht statt. Der Aufgabentyp „Vergleich“ wie auch der Aufgabentyp „Experimentieren“ war seltener und wurde nur bei einem Drittel der Befragten regelmäßig, bei einem weiteren Drittel ab und zu im Unterricht eingesetzt.

Abb. 9 zeigt die Verwendung der einzelnen Kontexte bzw. Aufgabentypen im Unterricht. Der Punkt verdeutlicht den Mittelwert, der Strich die Standardabweichung. Von den in der Hauptstudie angewandten Kontexten fand einzig das „Lösen von Brausetabletten“ selten im Unterricht statt. Diese einfache chemische Reaktion wird aber

oftmals bereits in der Primarstufe im Unterricht durchgeführt bzw. die Schülerinnen und Schüler kennen diese Reaktion aus dem Alltag.

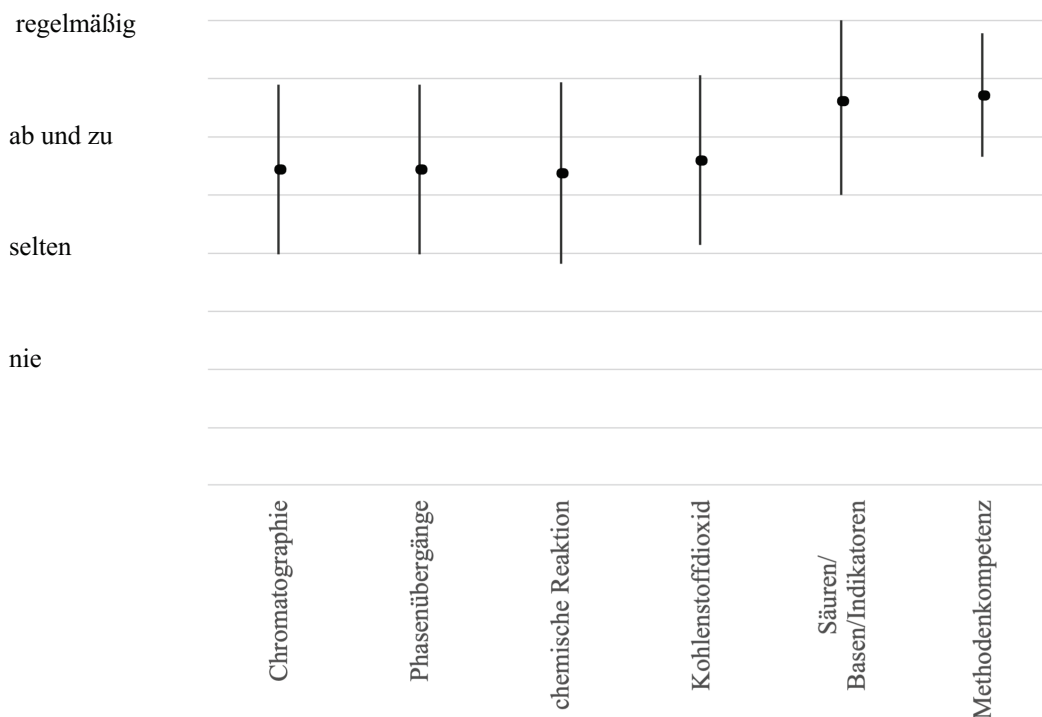


Abb. 9 ökologische Validierung

Die Tab. 45 zeigt den Mittelwert des Einsatzes im Unterricht. Zu dessen Ermittlung wurde eine Nennung von regelmäßig als 4, eine Nennung von ab und zu als 3, eine Nennung von selten als 2 und eine Nennung von nie als 1 gewertet. Ein Wert von über 3.0 im Mittelwert lässt auf eine Anwendung im Unterricht schließen.

Die Lehrpersonen der Sekundarstufe I wurden zudem befragt, ob die in der Tab. 45 aufgeführten Kontexte in ihren im Unterricht verwendeten Lehrmitteln enthalten sind und welches Lehrmittel sie im Unterricht einsetzen. Rund die Hälfte arbeitete mit selbsterarbeiteten Skripten, ergänzt durch Auszüge aus verschiedenen Lehrmitteln. Die übrigen setzen teilweise veraltete Lehrmittel, aber auch die neu erscheinenden Lehrmittel ein, welche zum Zeitpunkt der Erhebung noch nicht vollständig herausgegeben waren. Die Nennungen bezüglich der Inhalte der Lehrmittel decken sich mit den Nennungen bezüglich des Einsatzes im Unterricht.

Tab. 45 ökologische Validierung: Kontext, Aufgabentyp und Mittelwert

Kontext bzw. Aufgabentyp	Mittelwert
Chromatographie	3.10
Farbstift-Chromatographie	2.34
Phasenübergänge	2.97
Schmelzen von Eis	3.41
Einfache chemische Reaktionen	3.18
Lösen von Brausetabletten	2.31
CO ₂	3.45
Entstehung von CO ₂ mit Backpulver und Essig	2.00
Entstehung von CO ₂ mit Backpulver und Zitronensäure	2.97
Säuren und Basen	3.52
Indikatoren	3.10
Beobachtung von Phänomenen	3.83
Beobachtung Vergleich zweier Phänomene	3.03
Planung eines Experiments zur Erarbeitung von Zusammenhängen	3.21

7.4 Testgüte, Messinvarianz und Interpretation der Testwerte (F3)

Für die interne Validierung und die Beantwortung von F3 „Erfüllt das Messinstrument die Testgüte, die Messinvarianz und Generalisierbarkeit?“ (siehe Kapitel 6.7) wurden die Testgüte der Skala der Kompetenz des *wissenschaftlichen Skizzierens* (siehe Kapitel 7.4.1) und die Überprüfung Auftrennung der Itemschwierigkeiten sowohl kontext- wie aufgabentypabhängig (siehe Kapitel 7.4.2 - 7.4.4) untersucht. Außerdem wurde die Progression der Itemschwierigkeiten betrachtet (siehe Kapitel 7.4.5). Die Überprüfung der Messinvarianz und der daraus resultierenden Generalisierbarkeit (siehe Kapitel 7.7.4 und 7.4.6) erfolgte durch differenzierte Rasch-Analysen (DIF-Analysen).

7.4.1 Testgüte der Skala der Kompetenz des wissenschaftlichen Skizzierens

Zur Analyse der Items des Tests zur Kompetenz des *wissenschaftlichen Skizzierens* wurde ein personenzentriertes Rasch-Modell berechnet und die Testgüte ermittelt. Die Berechnung der Skala und ihrer Testgüte erfolgte eindimensional. Zur Analyse der

Aufgabenparameter wurden die gewichteten (Infit) und die ungewichteten (Outfit) Mean-Square-Fitstatistiken (MNSQ) ermittelt. Die Überprüfung erfolgte anhand eines T-Tests mit Hilfe z-standardisierter Werte. Werte zwischen 0.8 und 1.2 gelten für die Infit- und Outfit-Werte als gut (Wu et al., 2007). T-Werte sollten zwischen - 2.0 und 2.0 liegen (Wu et al., 2007). Als Signifikanztest des Zusammenhangs der Variablen wurde ein χ^2 -Test eingesetzt (Wu et al., 2007).

Tab. 46 Item-Kennwerte der neun Items des wissenschaftlichen Skizzierens der Rasch-Analyse (siehe Abb. 10 links)

N	376
Infit-Werte	0.94 < infit < 1.06
Outfit-Werte	0.93 < outfit < 1.07
T-Wert	- 0.7 < T < 0.8
EAP/PV Reliabilität	0.709
Item-Separation-Reliabilität	0.987
Variance	0.451
Geschätzte Parameter	28
Discrimination	0.54-0.65

Die Item-Separation-Reliabilität mit 0.987 bzw. 0.992 gibt die Zuverlässigkeit der Schätzung der Itemparameter an. Die Differenzierung zwischen leichten und schwereren Items ist sehr gut. Die EAP/PV Reliabilität, die angibt, wie gut zwischen den Merkmalsausprägungen der Personen unterschieden werden kann, liegt mit 0.709 bzw. 0.747 für ein offenes Setting in einem gut akzeptablen Bereich (Wu et al., 2007) dies, obwohl Testinstrumente mit offenem Antwortformat oftmals über eine geringere Personenreliabilität verfügen als Untersuchungen mit geschlossenen Antwortformaten (Stecher & Klein, 1996). Ähnliche Werte der Reliabilität erreichten auch andere Studien mit offenen Antwortformaten, wie bspw. im Projekt ExKoNawi (Bonetti et al., 2018).

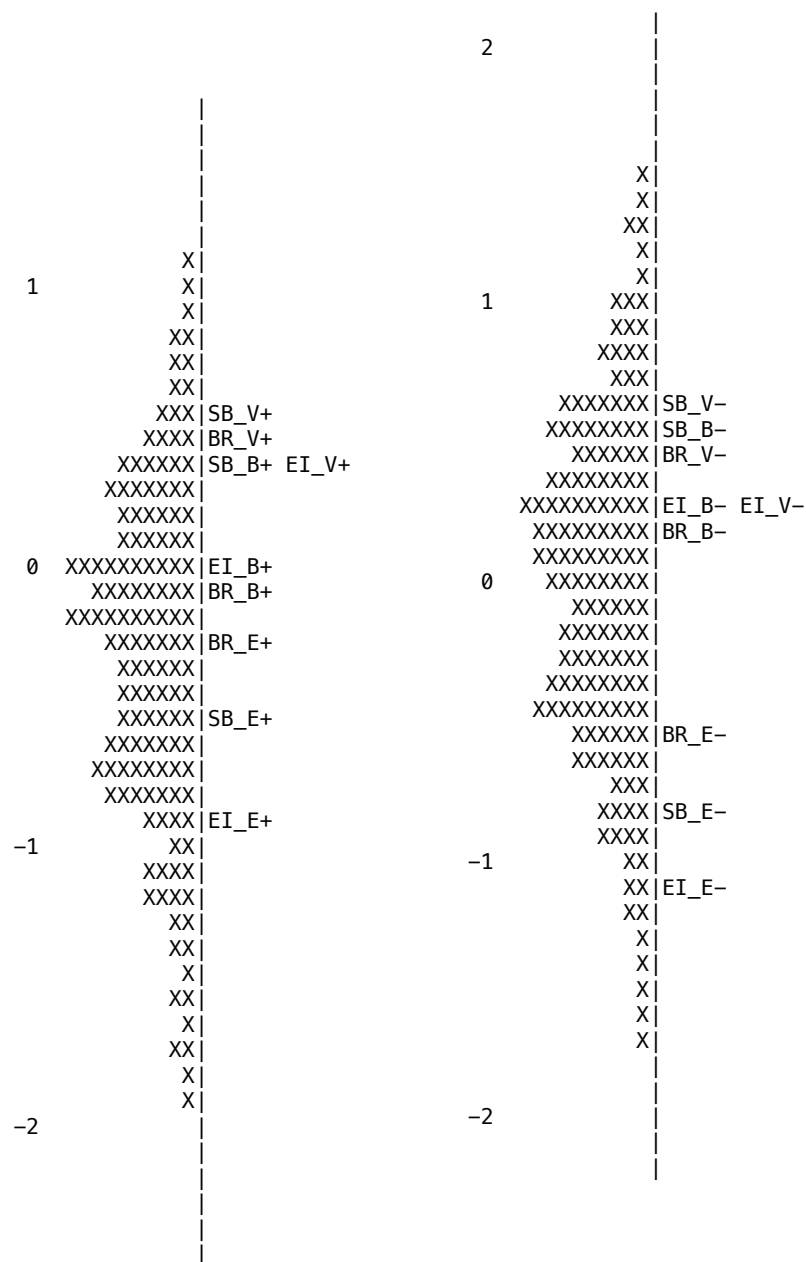


Abb. 10 Wright-Map der Itemschwierigkeit, aufgeteilt nach Aufgabentypen „Wissenschaftlichkeit der Skizze“ mit den aufgabentypspezifischen Kategorien links ($x = 2.4$ Fälle), ohne aufgabentypspezifische Kategorie rechts ($x = 2.6$ Fälle)

Die Rasch-Analyse ergab für die neun eingesetzten Items des *wissenschaftlichen Skizzierens* eine zufriedenstellende Modellpassung (siehe Tab. 46 und 47) (Wu et al., 2007). Die Wright-Map ist in der Abb. 10 links (Tab. 46) und in der Abb. 10 rechts (Tab. 47) abgebildet.

Tab. 47 Item-Kennwerte der neun Items des wissenschaftlichen Skizzierens der Rasch-Analyse (siehe Abb. 10 rechts)

N	376
Infit-Werte	$0.94 < \text{infit} < 1.10$
Outfit-Werte	$0.93 < \text{outfit} < 1.07$
T-Wert	$-0.7 < T < 1.3$
EAP/PV Reliabilität	0.747
Item-Separation-Reliabilität	0.992
Variance	0.491
Geschätzte Parameter	28
Discrimination	0.51-0.70

Zur Analyse der Schwierigkeit der Items ohne die aufgabentypenspezifischen Kategorien (siehe Kapitel 6.1 und 6.2 und Tab. 6) „Unterschied erfasst und Vergleich dargestellt“ und „Variablen erfasst und vollständig kontrolliert“, wie auch zur Analyse nach dem Zusammenhang der aufgabentypenspezifischen Kategorien, wurde deshalb einerseits nur die allgemein gültige Kategorie „*Wissenschaftlichkeit der Skizzen*“ (siehe Abb. 10 rechts) (Item-Bezeichnung ergänzt mit -), andererseits die Itemschwierigkeit mit den aufgabentypenspezifischen Kategorien (Item-Bezeichnung ergänzt mit +), (siehe Abb. 10 links) berücksichtigt. Zur Bezeichnung der Items gilt „B“ für den Aufgabentyp „Beobachtung“, „V“ für den Aufgabentyp „Vergleich“ und „E“ für den Aufgabentyp „Experimentieren“. Die Kontexte werden wie folgt abgekürzt: „BR“ für „Brausetablette“, „EI“ für „Eis schmelzen“ und „SB“ für „Säuren und Basen“.

Die Rasch-Analyse zeigte für beide Berechnungen eine zufriedenstellende Modellpassung auf (siehe Tab. 46 für Abb. 10 links und Tab. 47 für Abb. 10 rechts) (Wu et al., 2007). Wird nur die allgemeine Kategorie berücksichtigt, ist die Varianz leicht grösser (siehe Abb. 10 rechts und Tab. 47). Die Varianz, welche die Streuung der Items um den Mittelwert angibt, liegt mit 0.451 in einem mittleren Bereich. Die Breite der Kompetenzabdeckung ist nicht sehr groß, da die Schwierigkeitsgrade der Items eher ähnlich sind. Dies ist in der Wright-Map ersichtlich (siehe Abb. 10 links).

Die Abb. 10 zeigt die Wright-Map für die Skalierung der neun Items, bei der der Mittelwert der Itemschwierigkeiten auf null gesetzt wurde. Die Itemschwierigkeit wurde durch die Rasch-Analyse mit dem Programm ConQuest (Wu et al., 2007) ermittelt.

7.4.2 Schwierigkeitserzeugende Objektmerkmale

Durch Betrachtung der mittleren Itemschwierigkeit der einzelnen Items kann der Kompetenztest gesamthaft und unabhängig von der Kontextabhängigkeit analysiert werden, was somit zur inneren Validierung des Messinstruments beiträgt. Da Personen- und Itemparameter eine gemeinsame Skala haben, kann ermittelt werden, ob der Personen- oder Itemparameter größer ist. Die Personen werden in einer Wright-Map (siehe Abb. 10 und 11) auf der Logit-Skala durch Kreuze (x) links dargestellt. Die Items werden durch die Item-Kürzel rechts dargestellt. Im unteren Bereich der Wright-Map sind die einfacheren Items wie auch Personen mit geringen Personenfähigkeiten zu finden. Schwierigere Items und Personen mit höheren Personenfähigkeiten befinden sich im oberen Bereich. Liegt die Lösungswahrscheinlichkeit der Person für ein bestimmtes Item (Personenparameter) bei 50 %, ist auch der Itemparameter bei 50 %. Lernende mit einer Personenfähigkeit von 0.002 Logits lösen bspw. das Item EI_B (Aufgabentyp: Beobachtung/ Kontext: Eis schmelzen) (siehe Abb. 10 rechts) mit einer Wahrscheinlichkeit von 50 %. Items, die darunter liegen werden mit einer geringeren, Items, die darüber liegen, mit einer höheren Wahrscheinlichkeit gelöst. Die mittlere Itemschwierigkeit des Kompetenztests liegt bei $M = 0.000$ ($SD = 0.4783$, $Min = -0.852$, $Max = 0.523$). Die Itemschwierigkeit wurde auf schwierigkeitserzeugende Objektmerkmale untersucht. Die schwierigkeitserzeugenden Objektmerkmale lassen sich sowohl auf den Aufgabentyp als auch auf den Kontext beziehen. Aufgabentyp und Kontexte werden im Folgenden einzeln und zusammen, in Form einzelner Items, betrachtet (siehe Kapitel 7.4.3 und 7.4.5).

7.4.3 Aufgabentypabhängige Unterschiede der Itemschwierigkeiten

Durch die getrennte Betrachtung und Analyse der Itemschwierigkeiten kann die innere Validierung des Messinstruments analysiert werden: Der Aufgabentyp „Experimentieren“ zeigt generell die einfachste Itemschwierigkeit. Ebenfalls erkennbar ist, dass Vergleichsaufgaben als schwieriger einzustufen sind als Beobachtungsaufgaben. Dies steht im Widerspruch zu den Ergebnissen von Alfieri et al. (2013), die Vergleiche generell als einfacher einstufen (siehe Kapitel 3.1.2.2).

Wird nur die „*Wissenschaftlichkeit der Skizze*“ (siehe Abb. 10 rechts) berücksichtigt, fallen die Items zum Kontext „Eis schmelzen“ ähnlich schwierig aus. Bei den Kontexten „Brausetablette“ und „Säuren und Basen“ ist jedoch ersichtlich, dass die Itemschwierigkeit des Aufgabentyps „Vergleich“ als grösser einzustufen ist als beim Aufgabentyp „Beobachtung“, auch dann, wenn nur die allgemeinen Kategorien berücksichtigt werden (siehe Abb. 10 rechts). Dies kann darauf zurückgeführt werden, dass die Lernenden beim Aufgabentyp „Vergleich“ auf zwei Phänomene achten müssen. Oft ist feststellbar, dass die Lernenden zwar beide Phänomene darstellen, dies jedoch in nur *einem* Bild und nicht in einer Bildabfolge, wie so oft bei den Items des Aufgabentyps „Beobachtung“ (siehe Abb. 10).

Die beiden Bewertungen (siehe Abb. 10 links und rechts) unterscheiden sich auch darin, dass es zwischen den Aufgabentypen „Beobachtung“/„Vergleich“ und dem Aufgabentyp „Experimentieren“ einen unterschiedlich großen Gap gibt. Dieser Gap ist dann plausibel, wenn man bei den Aufgabentypen „Vergleich“ und „Experimentieren“ mehr Qualitätsansprüche anlegt (also die zusätzliche Kategorie berücksichtigt (siehe Abb. 10 links)), d.h. de facto, wenn man strenger kodiert, müsste sich der Gaps zwischen „Beobachtung“ und „Experimentieren“ und „Beobachtung“ und „Vergleich“ eigentlich schließen oder zumindest verringern. Da sich aber vor allem der Gap zwischen „Vergleich“ und „Beobachtung“ schließt bedeutet dies, dass die Zusatzkodierung den Lernenden hilft ein höheres Niveau zu erreichen. Eine mögliche Erklärung ist, dass der Aufgabentyp „Experimentieren“ im Bereich der „*Wissenschaftlichkeit der Skizze*“ sich stärker von den anderen beiden Aufgabentypen unterscheidet und dadurch einfacher erscheint.

7.4.4 Kontextabhängigkeit der Itemschwierigkeiten

Durch die getrennte Betrachtung der Itemschwierigkeiten kann die innere Validierung des Messinstruments analysiert werden: Die Kontexte weisen unterschiedliche Schwierigkeitsgrade auf, die durch die Itemschwierigkeiten erkennbar werden. Der Kontext „Säuren und Basen“ ist in den Aufgabentypen „Beobachtung“ (SB_B) und „Vergleich“ (SB_V) der schwierigste Kontext (siehe Abb. 10 rechts). Die Kodierung dieses Kontexts verlangt, im Gegensatz zu den anderen beiden Kontexten, eine Beschriftung der Chemikalien, damit die Eindeutigkeit vollständig erfüllt ist. Zudem zeigt sich das Phänomen durch Farbwechsel, die in der Skizze dargestellt werden müssen. Dies fällt den Lernenden offenbar schwerer, weil sie nicht mit Farben

arbeiten. Außerdem kann hier der Aspekt der Farbenblindheit eventuell eine Rolle spielen, welcher jeden zwölften Knaben betrifft, bei Mädchen mit nur einem Prozent jedoch selten auftritt. Die anderen beiden Kontexte fallen in ihrem Schwierigkeitsgrad ähnlich aus (siehe Abb. 10 rechts).

Während der Kontext „Säuren und Basen“ bei den Aufgabentypen „Beobachtung“ und „Vergleich“ der anspruchsvollste ist, scheint dies für den Aufgabentyp „Experimentieren“ beim Kontext „Brausetablette“ (BR_E) der Fall zu sein. Die Kontexte „Eis schmelzen“ und „Brausetablette“ scheinen im Schwierigkeitsgrad ähnlich eingestuft werden zu können.

Der Aufgabentyp „Vergleich“ verfügt zusätzlich über die Kategorie „Unterschied erfasst und Vergleich dargestellt“ (siehe Tab. 6) mit fünf bis sieben Indikatoren (siehe Tab. 38). Zwei Indikatoren stehen für die Darstellung des Vergleichs, einer für die Darstellung der Reaktion, die weiteren zwei bis vier für die Darstellung der Unterschiede der beiden Reaktionen oder Phänomene. Da nicht jeder Kontext die gleiche Anzahl an Unterschieden mitbringt, schwankt dieser Wert, folglich sind auch die Schwellenwerte (siehe Kapitel 6.5.2). leicht unterschiedlich.

Für das Niveau 3 des Kontexts „Säuren und Basen“ müssen alle fünf Indikatoren erfüllt sein. Dies kann ein Grund für den hohen Schwierigkeitsgrad dieses Niveaus sein (siehe Kapitel 6.5.2). Das Niveau 1 derselben Aufgabe ist hingegen tief hinsichtlich des Schwierigkeitsgrades angesiedelt. Ein Grund dafür kann der niedrige Schwellenwert zur Erfüllung dieses Niveaus sein (siehe Abb. 11).

Der Aufgabentyp „Experimentieren“ zeigt den niedrigsten Schwierigkeitsgrad (siehe Abb. 10 rechts). Alle drei Kontexte dieses Aufgabentyps weisen die gleiche Anzahl Indikatoren (siehe Anhang) und identische Schwellenwerte (siehe Kapitel 6.5.2) auf. Die Unterschiede im Schwierigkeitsgrad sind daher einzig kontextbedingt (siehe Tab. 19). Unterschiede in der Kodierung zeigen sich im Kontext „Säuren und Basen“, bei dem in der Facette „Eindeutigkeit“ (siehe Tab. 7) eine Legende oder Beschriftung verlangt wird, im Gegensatz zu den anderen beiden Kontexten. Dies hat beeinflusst die Erreichung der Niveaus 2 und 3.

7.4.5 Schwellenwerte der Itemschwierigkeiten

Zur Validierung der Progression kann die Struktur der Schwellenwerte (siehe Kapitel 6.5.2) analysiert werden: Inwiefern ist die Struktur über die Items verallgemeinerbar,

und wie gut sind einzelnen Progressionsstufen trennbar. Abb. 11 zeigt die Wright-Map mit den Schwellenwerten einer eindimensionalen Rasch-Analyse. Zur einfacheren Lesbarkeit wurden die Items der drei Aufgabentypen in verschiedenen Spalten aufgeführt. Daraus ist ersichtlich, wie schwierig es bspw. ist, den Schwellenwert 3 eines Items zu erreichen.

Die Rasch-Analyse ergab für die Berechnungen eine zufriedenstellende Modellpassung (siehe Tab. 48) (Wu et al., 2007).

Innerhalb eines Aufgabentyps gibt es keine Überlappungen der Schwellenwerte ungleicher Niveaus, was bedeutet, dass die Schwellen eindeutig die Niveaus differenzieren. Dies gilt innerhalb eines Aufgabentyps, jedoch nicht über die Aufgabentypen hinweg. Dieser Befund könnte gegen die Annahme eines eindimensionalen Konstrukts gedeutet werden (siehe Abb. 11).

Tab. 48 Item-Kennwerte der neun Items der Rasch-Analyse (Abb. 11)

N	376
Infit-Werte	$0.94 < \text{infit} < 1.04$
Outfit-Werte	$0.94 < \text{outfit} < 1.10$
T-Wert	$-0.7 < T < 0.8$
EAP/PV Reliabilität	0.709
Item-Separation-Reliabilität	0.987
Variance	0.451
Geschätzte Parameter	28
Discrimination	0.54-0.65

Das Niveau 1 des Items BR_V ist das schwierigste 1. Niveau aller Items. Da grundsätzlich Vergleiche leichter fallen als Beobachtungen, liegt die Begründung in der Kodierung. Für das Erreichen des Niveaus 1 müssen mehr Kriterien erfüllt sein – es gibt also eine größere Hürde für das Erreichen dieses Niveaus, da mehr Indikatoren kodiert werden. Bei den höheren Niveaus des Items BR_V ist diese Überhöhung weniger ausgeprägt vorhanden.

Die Schwellenwerte 1 und 2 des Kontexts „Brausetablette“ sind ähnlich niedrig wie der Schwellenwert des Kontextes „Eis schmelzen“. Alle Schwellenwerte des Aufga-

bentyps „Experimentieren“ des Kontextes „Eis schmelzen“ sind die niedrigsten. Der Kontext scheint weniger relevant zu sein, da die gleichen Kontexte ungefähr dieselben Schwellen haben. Die Kontexte strukturieren also weniger als die Aufgabentypen (siehe Abb. 11).

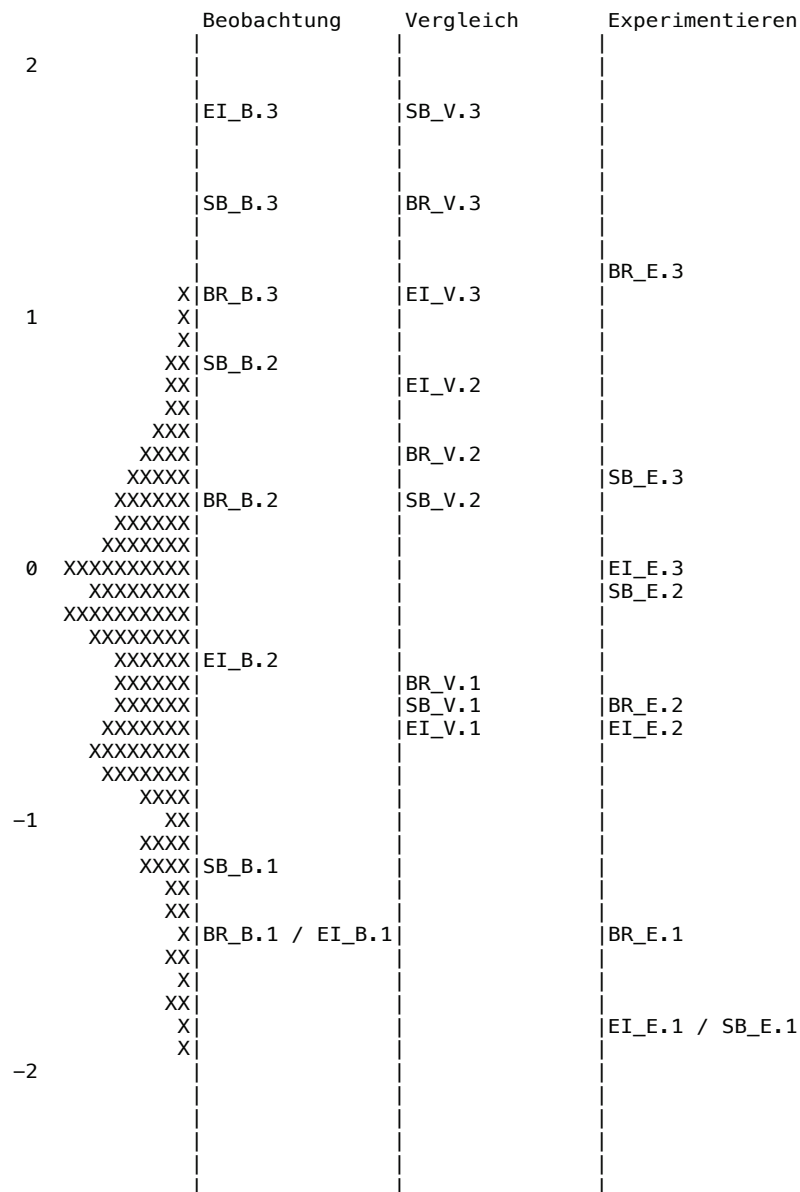


Abb. 11 Wright-Map der Itemschwierigkeiten (Schwellenwerte, getrennt nach Aufgabentyp und Item) ($x = 2.6$ Fälle)

Nachfolgend werden einige Items spezifisch beleuchtet und deren Itemschwierigkeit genauer analysiert. Die Auswahl ist als exemplarisch zu verstehen. Als Auswahlkri-

terien galten die höchsten Schwierigkeiten für das Erreichen des Niveaus 3 je Aufgabentyp (SB_V und BR_E) wie auch Items mit – bedingt durch schlechte Interrater-Übereinstimmungen – gestrichenen Indikatoren (EI_B und BR_V). Es handelt sich dabei um eine Besprechung der Schwellen und somit der Validierung.

7.4.5.1 Analyse Itemschwierigkeit BR_E

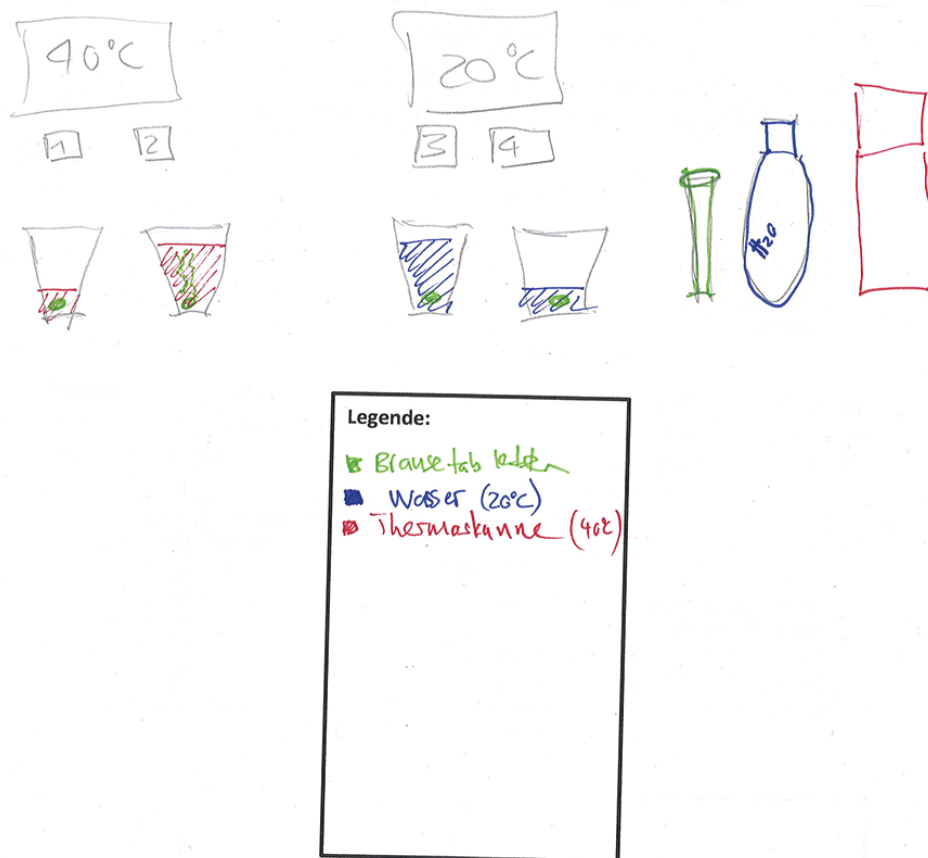


Abb. 12 Schülerskizze des Aufgabentyps „Experimentieren“ im Kontext „Brausetablette“ (Item BR_E)

Abb. 12 zeigt eine Schülerskizze des Aufgabentyps „Experimentieren“ im Kontext „Brausetablette“ (BR_E). Nahezu die Hälfte der Lernenden zeichnet die Reaktion der Brausetablette (Blasenentstehung) ein (siehe Abb. 12 zweites Gefäß). Bei den anderen beiden Kontexten ist die Anzahl der Lernenden, die die Reaktion darstellt, klein. Dieser Unterschied fließt in die Facette „Vollständigkeit“ ein und hat bereits ab dem Schwellenwert 1 Einfluss auf die Schwierigkeit der Aufgaben (siehe Abb. 10 und Kapitel 6.5.2). Vollumfänglich bemerkbar macht sich der Unterschied in der Schwierig-

keit erst bei der Schwelle zum Schwellenwert 3, da für diesen Schwellenwert alle drei Indikatoren der Facette „Vollständigkeit“ erfüllt sein müssen.

Der Schwellenwert 1 dieses Kontextes wie auch der Schwellenwert 1 des Kontextes „Säuren und Basen“ sind die am leichtesten zu erreichenden Schwellenwerte aller Items.

Alle drei Kontexte weisen eine zu kontrollierende Variable (Wasser- bzw. Teemenge) auf (siehe Anhang). Die zweiten zu kontrollierenden Variablen zeigen in den Kontexten „Eis schmelzen“ und „Brausetablette einen Unterschied bei der Substanz, in die etwas zugegeben wird (Salz- und Süßwasser bzw. unterschiedliche Wassertemperaturen). Beim Kontext „Säuren und Basen“ handelt es sich bei der zweiten Variablen um unterschiedliche Substanzen, die zugegeben werden. Bei der Betrachtung der Mittelwerte (siehe Anhang) scheint diese Variablenkontrolle des Kontexts „Säuren und Basen“ den Lernenden schwerer zu fallen, da nur gut die Hälfte diesen Indikator erfüllt. Sowohl die Bestimmung der Variablen als auch die Beschriftung beim Kontext „Säuren und Basen“ könnten die Erhöhung des Schwierigkeitsgrades erklären.

7.4.5.2 Analyse Itemschwierigkeit Item EI_B

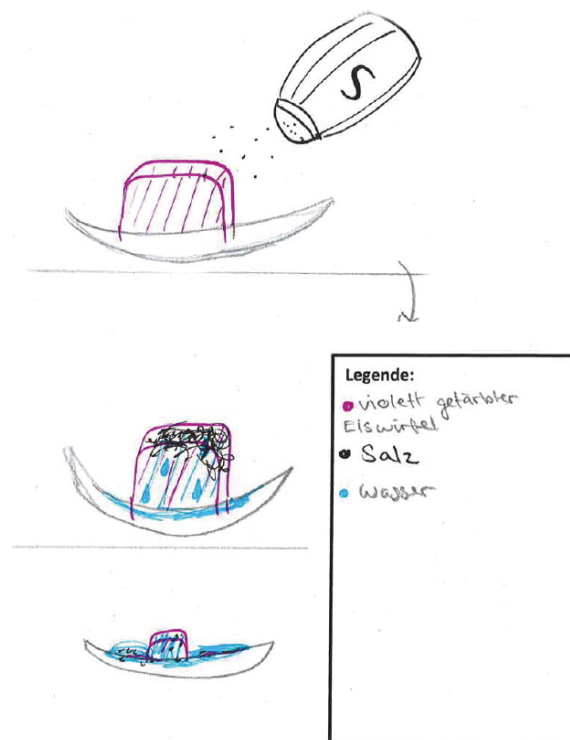


Abb. 13 Schülerskizze des Aufgabentyps „Beobachtung“ im Kontext „Eis schmelzen“ (Item EI_B)

Der Schwellenwert 3 des Aufgabentyps „Beobachtung“ im Kontext „Eis schmelzen“ (EI_B) (siehe Abb. 13) ist am schwierigsten zu erreichen (siehe Abb. 11). Diese Aufgabe verfügt, bedingt durch die Streichung eines Indikators (siehe Kapitel 6.6.4), in der Facette „fachliche Korrektheit“ über nur einen Indikator, und dieser muss für das Erreichen des Schwellenwerts 3 erfüllt sein. Alle anderen Aufgabentypen verfügen über mindestens zwei Indikatoren für diese Facette, wobei nur ein Indikator für die Erfüllung des Schwellenwerts 3 erfüllt sein muss (siehe Kapitel 6.5.2).

7.4.5.3 Analyse Itemschwierigkeit Item SB_V

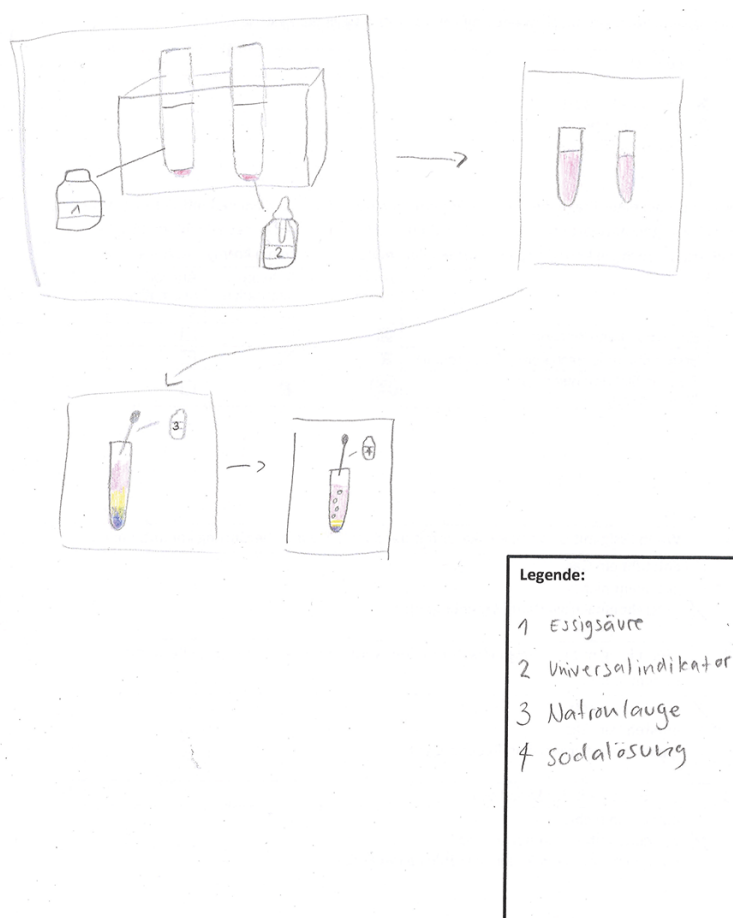


Abb. 14 Schülerskizze des Aufgabentyps „Vergleich“ im Kontext „Säuren und Basen“ (Item SB_V)

Der Schwellenwert 3 des Aufgabentyps „Vergleich“ im Kontext „Säuren und Basen“ (SB_V) (siehe Abb. 14) ist als ebenso anspruchsvoll eingestuft (siehe Abb. 11). Der hohe Schwierigkeitsgrad kann einerseits durch die Notwendigkeit der Beschriftung der Chemikalien in der Skizze erklärt werden wie auch durch den weniger vorhandenen Alltagsbezug dieses Items. Ein weiterer Grund ist der Schwellenwert der Kategorie „Unterschied erfasst und Vergleich dargestellt“: Für die Erreichung des Schwel-

lenwerts 3 müssen alle fünf Indikatoren dieser Kategorie erfüllt sein (siehe Kapitel 6.5.2).

Ebenfalls einen hohen Schwierigkeitsgrad weisen im Aufgabentyp „Beobachtung“ der Kontext „Säuren und Basen“ (SB_B) und im Aufgabentyp „Vergleich“ der Kontext „Brausetablette“ (SB_V) auf (siehe Abb. 14).

7.4.5.4 Analyse Itemschwierigkeit Item BR_V

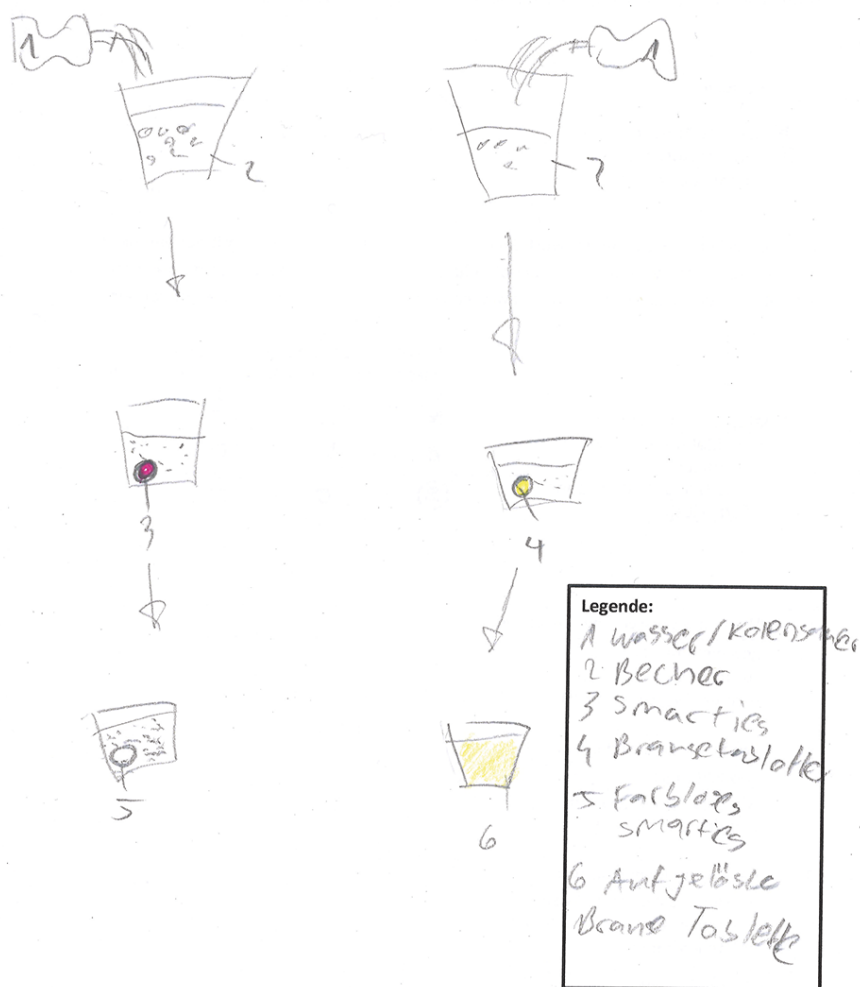


Abb. 15 Schülerskizze des Aufgabentyps „Vergleich“ im Kontext „Brausetablette“ (Item BR_V)

Abb. 15 zeigt eine Schülerskizze des Aufgabentyps „Vergleich“ im Kontext „Brausetablette“ (BR_V). Alle drei Schwellenwerte dieses Items sind hoch (siehe Abb. 11). Die Schwellenwerte sind den Schwellenwerten der anderen beiden Kontexte dieses Aufgabentyps ähnlich, auch wenn die Anzahl der Indikatoren bei dieser Aufgabe, bedingt durch schlechte Interrater-Übereinstimmungen, reduziert werden musste (siehe

Kapitel 6.6.4). Auch in der Kategorie „Vergleich erfasst und Unterschied dargestellt“ musste für diese Aufgaben ein Indikator wegen unzureichender Interrater-Übereinstimmung ausgeschlossen werden (siehe Kapitel 6.6.4). Für die Erreichung des Schwellenwerts 3 sind fünf von sechs Indikatoren notwendig. Im Vergleich zu den anderen beiden Kontexten dieses Aufgabentyps sind viele Unterschiede vorhanden, die erkannt werden müssen, was die Aufgaben schwieriger macht.

7.4.5.5 Analyse Itemschwierigkeit Item SB_B

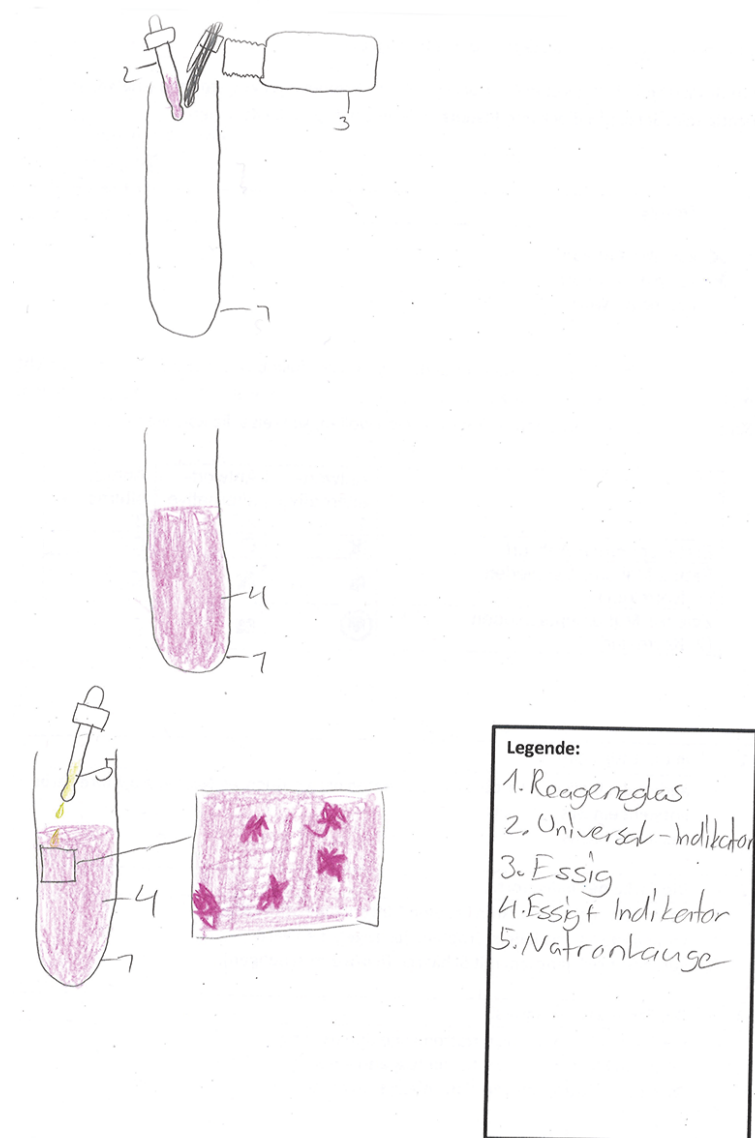


Abb. 16 Schülerskizze des Aufgabentyps „Beobachtung“ im Kontext „Säuren und Basen“ (Item SB_B)

Das Item im Aufgabentyp „Beobachtung“ des Kontextes „Säuren und Basen“ (SB_B) (siehe Abb. 16) hat den höchsten Schwierigkeitsgrad für diesen Aufgabentyp (siehe Abb. 11). Alle Schwellenwerte liegen hier im Schwierigkeitsgrad höher als bei den anderen Kontexten. Kontextspezifische Objektmerkmale und Schwierigkeiten (Farbänderungen, die darzustellen sind, notwendige Beschriftung, Vorhanden-Sein des Reagenzglas-Ständers in der Video-Vignette) erhöhen den Schwierigkeitsgrad dieses Items gegenüber den anderen beiden Kontexten.

Aus der Wright-Map (siehe Abb. 11) wird ersichtlich, dass der Schwellenwert 1 des Aufgabentyps „Experimentieren“ den niedrigsten Schwierigkeitsgrad aufweist. Ebenso ist ersichtlich, dass der Schwellenwert 1 des Aufgabentyps „Vergleich“ deutlich schwieriger zu erfüllen ist als der Schwellenwert 1 des Aufgabentyps „Beobachtung“. Die Schwellenwerte zur Erreichung des Niveaus 1 wie auch die Indikatoren zu deren Erfüllung sind gleich. Die höhere Schwierigkeit des Aufgabentyps „Vergleich“ ist vor allem dadurch bedingt, dass bei diesem Aufgabentyp zwei Phänomene miteinander verglichen werden, also beide beobachtet werden müssen. Bei der Erfüllung des Schwellenwerts 2 und des Schwellenwerts 3 ist eine Differenz in den Aufgabentypen nicht mehr feststellbar. Objektmerkmale wie auch damit verbundene Kodier- unterschiede fallen dabei mehr ins Gewicht.

7.4.6 Messinvarianz des Messinstruments

Messinvarianz ist der umfassende Begriff, der genutzt wird, wenn ein Instrument unabhängig von Ort, Sample und Zeitpunkt gleich misst. Ein Instrument ist messinvariant, wenn dieselben Items für unterschiedliche Lernenden in derselben Schwierigkeitshierarchie aufeinander folgen.

Die Überprüfung der Messinvarianz wurde bezüglich Personenmerkmale, Geschlecht, Jahrgangsstufe, Klassenniveau, Anzahl gesprochener Sprachen und Abhängigkeit der Personenfähigkeit mit Lösewahrscheinlichkeit der Items vorgenommen. Mit DIF-Analysen wurde betrachtet, ob die konkreten Leistungen für eine willkürlich gewählte Paarung unterschiedlich ausgeprägt waren. Ermittelt wurde, wie stark sich die einzelnen Items bezüglich eines Personenmerkmals unterschieden. Eine Differenz in den geschätzten Itemschwierigkeiten unter 0.213 liegt laut Santelices und Wilson (2012) und Paek (2004) im vernachlässigbaren Bereich. Die Abhängigkeiten zur Kompetenzausprägung des *wissenschaftlichen Skizzierens* wurden mittels Differen-

tial-Item-Functioning (DIF) für mehrkategorielle Partial Credits Items (siehe Kapitel 6.8.2) einer eindimensionalen Rasch-Analyse ermittelt (Wu et al., 2007, S. 85). Dabei konnten auch Aussagen über die Leistung der Stichprobe bezüglich des differenzierten Personenmerkmals gemacht werden.

7.4.6.1 Messinvarianz bezüglich des Geschlechts

Die Messinvarianz bezüglich des Geschlechts (48.4 % weibliche Lernende) wurde mittels einer Differential-Item-Functioning (DIF) für mehrkategorielle Partial Credits Items einer eindimensionalen Rasch-Analyse überprüft (Wu et al., 2007, S. 85). Es wurde das Modell „item-geschlecht+item*geschlecht+item*step“ gerechnet (Wu et al., 2007, S. 85). Alle neun untersuchten Items weisen eine Differenz bezüglich des Geschlechts in den Itemschwierigkeiten von unter 0.213 auf, was laut Santelices und Wilson (2012) und Paek (2004) im vernachlässigbaren Bereich liegt.

Tab. 49 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich des Geschlechts

N	376	Item	Mädchen	Jungen
Infit-Werte	0.92 < infit < 1.06	BR_B	0.036	-0.036
Outfit-Werte	0.93 < outfit < 1.08	EI_B	0.055	-0.055
T-Wert	- 0.5 < T < 0.7	SB_B	0.018	-0.018
EAP/PV Reliabilität	0.640	BR_V	0.039	-0.039
Item-Separation-Reliabilität	0.987	EI_V	0.084	-0.084
Variance	0.457	SB_V	-0.070	0.070
		BR_Z	0.082	-0.082
		EI_Z	-0.086	0.086
		SB_Z	-0.158	0.158

Die größte Differenz in der nach dem Geschlecht differenzierten Itemschwierigkeit weist mit 0.158 der Aufgabentyp „Experimentieren“ im Kontext „Säuren und Basen“ (SB_E) auf. Es besteht kein DIF-Effekt auf der Ebene der Items, der Test ist also messinvariant bezüglich des Geschlechts. Das heißt, der Test zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* weist keinen Geschlechterunterschied auf (siehe Tab. 49). Außerdem zeigt der Gesamttest auch keinen signifikanten Leistungsunterschied in den Personenfähigkeiten zwischen weiblichen Lernenden (M=-0.006, SD=0.026, N=193) und männlichen Lernenden (M=0.006, SD= 0.026, N=183). Die Rasch-Analyse zeigte für die neun eingesetzten Items eine zufriedenstellende Modellpassung auf (siehe Tab. 49) (Wu et al., 2007).

7.4.6.2 Messinvarianz bezüglich der Jahrgangsstufe

Tab. 50 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich der Jahrgangsstufe

N	376	Item	7. Jahrgangsstufe	9. Jahrgangsstufe
Infit-Werte	$0.92 < \text{infit} < 1.07$	BR_B	-0.151	0.151
Outfit-Werte	$0.92 < \text{outfit} < 1.07$	EI_B	-0.042	0.042
T-Wert	$-0.7 < T < 0.6$	SB_B	0.058	-0.058
EAP/PV Reliabilität	0.987	BR_V	0.118	-0.118
Item-Separation-Reliabilität	0.669	EI_V	-0.016	0.016
Variance	0.441	SB_V	0.063	-0.063
		BR_E	-0.002	0.002
		EI_Z	-0.003	0.003
		SB_E	-0.031	0.031

Die Messinvarianz bezüglich Jahrgangsstufe (54.0 % 7. Jahrgangsstufe) wurde mittels einer Differential-Item-Functioning (DIF) für mehrkategoriale Partial Credits Items einer eindimensionalen Rasch-Analyse überprüft (Wu et al., 2007, S. 85). Es wurde das Modell „item - jahrgangsstufe + item*jahrgangsstufe + item*step“ gerechnet (Wu et al., 2007, S. 85).

Alle neun untersuchten Items weisen eine Differenz bezüglich der Jahrgangsstufe in den Itemschwierigkeiten von unter 0.213 auf, was laut Santelices und Wilson (2012) und Paek (2004) im vernachlässigbaren Bereich liegt. Die größten Differenzen in der Itemschwierigkeit, differenziert nach der Jahrgangsstufe, weisen mit 0.151 bzw. 0.118 die Aufgabentypen „Beobachtung“ und „Vergleich“ im Kontext „Brausetablette“ (BR_B und BR_V) auf. Diese liegen beide jedoch im vernachlässigbaren Bereich (Paek, 2004; Santelices & Wilson, 2012). Die Items sind messinvariant zwischen den Jahrgangsstufen, dies heißt, es gibt kein Item, welches einen Vorteil für Lernende der 7. oder 9. Jahrgangsstufe bietet. Es zeigt sich auch kein DIF-Effekt auf der Ebene der Items (siehe Tab. 50).

Der Test zeigt Leistungsunterschiede in den Personenfähigkeiten zwischen den Lernenden der 7. Jahrgangsstufe ($M=-0.106$, $SD=0.025$, $N=203$) und den Lernenden der 9. Jahrgangsstufe ($M=0.106$, $SD=0.025$, $N=173$) welche hoch signifikant sind. Für die Stichprobe der Lernenden der 9. Jahrgangsstufe ist das ganze Aufgabenset leichter

als für die Stichprobe der Lernenden der 7. Jahrgangsstufe. Die Rasch-Analyse zeigte für die neun eingesetzten Items eine ausreichende zufriedenstellende Modellpassung auf (siehe Tab. 50) (Wu et al., 2007).

7.4.6.3 *Messinvarianz bezüglich der Anzahl der gesprochenen Sprachen*

Die Messinvarianz bezüglich der Anzahl der gesprochenen Sprachen wurde mittels einer Differential-Item-Functioning (DIF) für mehrkategoriale Partial Credits Items einer eindimensionalen Rasch-Analyse überprüft (Wu et al., 2007, S. 85). Es wurde das Modell „item - sprache + item*sprache + item*step“ gerechnet (Wu et al., 2007, S. 85). Erfasst wurde dabei, ob die Lernenden nur Deutsch (40.9 %) (Schweizerdeutsch und/oder Schriftsprache) oder eine (47.8 %) oder mehr als eine (11.3 %) Fremdsprache zusätzlich zu Deutsch zu Hause sprechen. Diejenigen Lernenden, die nur Deutsch sprechen, sind in der Tab. 51 unter Deutsch erfasst. Die Lernenden mit 1 bzw. mind. 2 Fremdsprachen (siehe Tab. 51) sprechen zusätzlich zu Deutsch, das sie mindestens in der Schule sprechen, noch 1 bzw. mindestens 2 Fremdsprachen zuhause.

Mit einer Differenz von 0.714 in den geschätzten Itemschwierigkeiten zwischen Schülerinnen und Schülern, die nur Deutsch und solchen, die zwei Fremdsprachen sprechen, liegt der Aufgabentyp „Experimentieren“ im Kontext „Eis schmelzen“ (EI_Z) im großen Bereich von >0.319 (Paek, 2004; Santelices & Wilson, 2012). Eine mögliche inhaltliche Begründung kann aus curricularer Sicht wie auch aus sprachlicher Sicht des Item-Textes nicht vorliegen, darum handelt es sich höchstwahrscheinlich um ein Artefakt. Daraufhin deutet auch, dass die Differenzen der geschätzten Itemschwierigkeiten für dieses Item zwischen Lernenden, die eine Fremdsprache sprechen zu denjenigen, die nur Deutsch sprechen, mit 0.177 vernachlässigbar ist (Paek, 2004; Santelices & Wilson, 2012). Für alle anderen Items liegen die Differenzen der geschätzten Itemschwierigkeiten zwischen allen drei Gruppen im vernachlässigbaren Bereich (Paek, 2004; Santelices & Wilson, 2012). Die Items sind messinvariant - abgesehen vom Aufgabentyp „Experimentieren“ im Kontext „Eis schmelzen“ (EI_Z) in Bezug auf die Anzahl Sprachen, die die Lernenden sprechen. Einzig dieses Item bietet einen Vorteil für Lernende, die nicht mehrere Fremdsprachen sprechen (siehe Tab. 51).

Tab. 51 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich der Anzahl der gesprochenen Sprachen

N	376	Item	Deutsch	1 Fremd- sprache	mind. 2 Fremd- sprachen
Infit-Werte	$1.01 < \text{infit} < 1.05$	BR_B	-0.194	-0.024	0.217
Outfit-Werte	$0.99 < \text{outfit} < 1.10$	EI_B	-0.020	-0.119	0.139
T-Wert	$0.0 < T < 0.7$	SB_B	-0.163	0.030	0.133
EAP/PV Reliabilität	0.974	BR_V	-0.024	0.043	-0.119
Item-Separation- Reliabilität	0.565	EI_V	-0.07	0.115	-0.122
Variance	0.409	SB_V	-0.081	-0.130	0.211
		BR_E	0.080	0.073	-0.153
		EI_Z	0.297	0.120	-0.417
		SB_E	0.098	-0.109	0.010

Der Test zeigt Leistungsunterschiede aufgrund der Anzahl gesprochener Sprachen. Der Leistungsunterschied in den Personenfähigkeiten zwischen Lernenden, die nur Deutsch sprechen ($M=0.201$, 0.035 , $N=149$), und Lernenden, die mindestens eine Fremdsprache sprechen, ist signifikant (Wu et al., 2007). Hingegen zeigen sich keine signifikanten Leistungsunterschiede zwischen Lernenden, die eine Fremdsprache ($M=-0.108$, 0.034 , $N=172$) oder mehrere Fremdsprachen ($M=-0.093$, 0.049 , $N=38$) sprechen (Wu et al., 2007). Für Lernende, die nur Deutsch sprechen, ist das Aufgabenset deshalb als Ganzes einfacher einzustufen. Diese Ergebnisse sind konform mit bekannten Ergebnissen und bestätigen somit die Annahme. Die Rasch-Analyse wies für die neun eingesetzten Items eine zufriedenstellende Modellpassung auf (siehe Tab. 51) (Wu et al., 2007).

7.4.6.4 Messinvarianz bezüglich des Leistungsniveaus

Die Messinvarianz bezüglich der Leistungsniveaus wurde überprüft. Unterschieden gab es zwischen Lernenden, die eine Sekundarschule besuchen, und Lernenden, die das Gymnasium besuchen. Die Lernenden, die das Gymnasium besuchen, waren zum Zeitpunkt der Erhebung alle in der 9. Jahrgangsstufe. Die Schülerinnen und Schüler der Sekundarschule besuchten entweder die 7. Oder die 9. Jahrgangsstufe und waren in unterschiedliche Niveaus der Sekundarschule eingeteilt, dies sowohl in niveau-

durchmischten wie auch nicht in niveaudurchmischten Klassen. Es wurde das Modell „item - niveau + item*niveau+ item*step“ gerechnet (Wu et al., 2007, S. 85).

Tab. 52 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich des Leistungsniveaus

N	376	Item	Sek I	Gymnasium
Infit-Werte	0.91 < infit < 1.07	BR_B	-0.396	0.396
Outfit-Werte	0.90 < outfit < 1.11	EI_B	-0.141	0.141
T-Wert	- 0.7 < T < 0.9	SB_B	0.206	-0.206
EAP/PV Reliabilität	0.984	BR_V	0.263	-0.263
Item-Separation-Reliabilität	0.720	EI_V	0.024	-0.024
Variance	0.435	SB_V	0.171	-0.171
		BR_E	0.017	-0.017
		EI_Z	-0.048	0.048
		SB_E	-0.096	0.096

Für den Aufgabentyp „Beobachtung“ im Kontext „Brausetablette“ (BR_B) ist die Differenz in den geschätzten Itemschwierigkeiten zwischen Lernenden der Sekundarstufe (N=232) und Lernenden des Gymnasiums (N=15) mit 0.396 groß (> 0.319) (Paek, 2004; Santelices & Wilson, 2012). Für den Aufgabentyp „Vergleich“ im Kontext „Brausetablette“ (BR_V) (N(Gymnasium)=33, N(Sekundarschule)=212) liegt die Differenz in den geschätzten Itemschwierigkeiten zwischen Lernenden der Sekundarschule und Lernenden des Gymnasiums mit 0.263 im moderaten Bereich (Paek, 2004; Santelices & Wilson, 2012). Bei allen anderen Items ist die Differenz der geschätzten Itemschwierigkeiten mit unter 0.213 vernachlässigbar (Paek, 2004; Santelices & Wilson, 2012) (siehe Tab. 52). Die beiden Items (BR_B und BR_V) mit DIF-Effekten zeigen beide gute Fitwerte (siehe Tab. 52) und sind Items mittlerer Schwierigkeit (siehe Abb. 10). Gründe könnten einerseits in der Alltäglichkeit des Kontextes, beim Hintergrundwissen, das im bildungsnäheren Umfeld ausgeprägter vorhanden ist, liegen. Andererseits könnte aber auch die Wahrnehmung der Blasenbildung des Lösens der Brausetablette bzw. des Smarties unterschiedlich fokussiert worden sind. Dies sollte aber mit der fokussierten Aufgabenstellung weitestgehend ausgeschlossen worden sein.

Zusammenfassend kann gesagt werden: Es liegt keine relevante Bevorzugung oder Benachteiligung einer der beiden Gruppen vor, sondern nur stichpunktartige DIF-Effekte auf der Ebene von zwei Items. Es liegen keine systematischen DIF-Effekte bezüglich des Leistungsniveaus der Lernenden vor. Ebenso wenig weisen die Lehrpläne bezüglich der beiden Items relevante Unterschiede auf, sodass nicht davon ausgegangen werden kann, dass der Test bestehende Unterschiede messen könnte. Die Differenzen im Aufgabentyp „Vergleich“ im Kontext „Brausetablette“ (BR_V) in den geschätzten Itemschwierigkeiten scheinen unbedenklich. Eine mögliche Erklärung ist die kleinen Stichproben der Gymnasiasten, woraus Artefakte resultieren könnten.

Der Test zeigt Leistungsunterschiede in den Personenfähigkeiten aufgrund des besuchten Schul-Niveaus. Der Leistungsunterschied zwischen den Lernenden der Sekundarschule ($M=0.235$, 0.026 , $N=328$) und den Gymnasiasten ($M=-0.235$, 0.026 , $N=47$) ist höchst signifikant (Wu et al., 2007). Lernenden des Gymnasiums fällt der Test leichter als Lernenden der Sekundarstufe I. Diese Ergebnisse sind konform mit bekannten Ergebnissen und bestätigen somit die Annahme. Die Rasch-Analyse zeigte für die neun eingesetzten Items eine zufriedenstellende Modellpassung auf (siehe Tab. 52) (Wu et al., 2007).

7.4.6.5 *Messinvarianz bezüglich der Personenfähigkeit*

Die Messinvarianz bezüglich der Personenfähigkeit wurde überprüft. Es wurde das Modell „item - faehigkeit + item*faehigkeit + step*item“ gerechnet (Wu et al., 2007, S. 85). Der Median der Personenfähigkeit aller Items des *wissenschaftlichen Skizzierens* wurde ermittelt und Lernenden mit einer Personenfähigkeit unter dem Medianwert das Prädikat „schwach“ (46.3 %), Lernenden mit einer Personenfähigkeit über dem Medianwert das Prädikat „stark“ (53.7 %) zugeordnet.

Der Leistungsunterschied in den Personenfähigkeiten zwischen den „starken“ Lernenden ($M=-0.691$, 0.026 , $N=201$) und den „schwachen“ Lernenden ($M=0.691$, 0.026 , $N=175$) ist höchst signifikant (Wu et al., 2007). Die Rasch-Analyse zeigte für die neun eingesetzten Items eine zufriedenstellende Modellpassung auf (siehe Tab. 53) (Wu et al., 2007).

Für den Aufgabentyp „Vergleich“ im Kontext „Brausetablette“ (BR_V) ist die Differenz in den geschätzten Itemschwierigkeiten zwischen „starken“ und „schwachen“ Lernenden mit 0.300 moderat (0.219 - 0.319) (Paek, 2004; Santelices & Wilson,

2012). Diese Aufgabe ist im Vergleich zu den restlichen Aufgaben für die besseren Lernenden auf der gleichen Logit-Skala einfacher als für die schwachen Lernenden. Bei allen anderen Items liegen die Differenzen der geschätzten Itemschwierigkeiten im vernachlässigbaren Bereich (Paek, 2004; Santelices & Wilson, 2012). Es kann von keiner Benachteiligung oder Bevorzugung einer der beiden Gruppen ausgegangen werden (siehe Tab. 53). Dadurch kann auch eine mögliche Bevorzugung durch die Leistungsniveaus als unfundiert betrachtet werden.

Tab. 53 Item-Kennwerte und DIF-Werte der neun Items der DIF-Analyse bezüglich der Personenfähigkeit

N	376	Item	Schwach	stark
Infit-Werte	$1.00 < \text{infit} < 1.07$	BR_B	-0.061	0.061
Outfit-Werte	$1.00 < \text{outfit} < 1.07$	EI_B	-0.062	0.062
T-Wert	$-0.1 < T < 0.5$	SB_B	0.073	-0.073
EAP/PV Reliabilität	0.989	BR_V	0.300	-0.300
Item-Separation-Reliabilität	0.016	EI_V	-0.008	0.008
Variance	0.005	SB_V	0.097	-0.097
		BR_E	0.013	-0.013
		EI_Z	-0.182	0.182
		SB_E	-0.169	0.169

7.5 Externe Validierung (F4-F6)

Beim Test zur Erfassung des Strategiewissens zum Experimentieren (NAW) wurden neun Items erhoben, im Mittel wurden 4.57 davon korrekt beantwortet (siehe Tab. 54). Beim Test zur Erfassung der Kognition (KFT) wurde die nonverbale Skala N3 erhoben, die aus fünfzehn Items besteht. Im Mittel wurden 8.74 korrekt beantwortet (siehe Tab. 54), was leicht unter den Mittelwerten der zu erwartenden Leistungen liegt. Für die 7. Jahrgangsstufe wäre mit einem Mittelwert von 10.81 (Standardabweichung 3.21) und für die 9. Jahrgangsstufe mit einem Mittelwert von 10.54 (Standardabweichung 3.57) zu rechnen (Heller & Perleth, 2000). Eine mögliche Erklärung für die schwachen Leistungen in dieser Studie könnten Ermüdungserscheinungen sein, da die Erfassung der externen Variablen im Anschluss an die Erhebung des *wissenschaftlichen Skizzierens* stattfand.

Die Leistungen für die Lesefähigkeit wurden in drei Subskalen aufgespalten: das Leseverständnis (LV), die Lesegenauigkeit (LGN) und die Lesegeschwindigkeit (LGS).

Die Leistungen für das Leseverstehen (LV) liegen mit einem T-Wert von 27.68, gemittelt über alle Schülerinnen und Schüler der Studie, leicht über den Werten für Lernende der 7. Jahrgangsstufe (23.9), jedoch unter den Leistungen für Lernende der 9. Jahrgangsstufe (nicht-gymnasial: 30.5, gymnasial: 39.0) (Schneider et al., 2007). Schneider et al. (2007) weisen zudem auf die niedrigeren Korrelationen zwischen KFT und LGVT hin (LV: 0.17, LGS: 0.14, LGN: 0.19).

Tab. 54 Resultate externe Variablen

	NAW	KFT	LV	LGN	LGS
Mittelwert	4.57	8.74	27.68	37.22	39.97
mögliche Werte	0-9	0-15	27-73	27-73	27-73
Standard-Abweichung	2.32	3.54	12.09	17.17	11.52

Zur Beantwortung der F4 und F6 (siehe Kapitel 5) wurden sowohl eine Regressionsanalyse wie auch ein Strukturgleichungsmodell angewandt. Die Resultate und deren Interpretation werden im Folgenden getrennt nach den beiden Analysemethoden aufgeführt (siehe Kapitel 7.5.1 und 7.5.3). Für die Beantwortung der F5 (siehe Kapitel 5) wurde neben einer prozentualen Verteilung aufgrund von unterschiedlicher Performanz ein Kendall-Tau-Test eingesetzt (siehe Kapitel 7.5.2).

7.5.1 Kognition, Strategiewissen zum Experimentieren und Lesefähigkeit (F4)

Zur Untersuchung, ob die Kognition (KFT), die Lesefähigkeit (LGVT) und das Strategiewissen zum Experimentieren (NAW) prädiktiv für die Personenfähigkeit des *wissenschaftlichen Skizzierens* sind (siehe F4), wurden Regressionsanalysen durchgeführt, die diese Zusammenhänge untersuchten. Die Personenfähigkeit des *wissenschaftlichen Skizzierens* wird als abhängige Variable (Kriteriums-Variable) eingesetzt und die Werte der Kognition, der Lesefähigkeit und des Strategiewissen zum Experimentieren VKS als Prädiktoren (unabhängige Variablen). Weder die Lesegeschwindigkeit noch die Lesegenauigkeit oder das Leseverständnis zeigten nach einer Korrelationsanalyse nach Pearson einen Zusammenhang mit der Personenfähigkeit des *wissenschaftlichen Skizzierens* (siehe Tab. 55). Die Kognition und das Strategiewissen zum Experimentieren zeigten einen bedeutsamen Zusammenhang zum *wissenschaftlichen Skizzieren* (siehe Tab. 56).

Tab. 55 Korrelationskoeffizienten zwischen der Personenfähigkeit des wissenschaftlichen Skizzierens und der Lesefähigkeit

Personenfähigkeit	Leseverständnis	0.119**
	Lesegeschwindigkeit	-0.097**
	Lesegenauigkeit	0.190*
(* $p \leq 0.01$, ** $p \leq 0.05$)		

Die schriftlich gestellten Aufgaben zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* stellen an die Lernenden Leseanforderungen (Gut, 2012) in einer fachlich korrekten, einfach verständlichen Sprache mit geringem Textumfang (siehe Kapitel 6.3.1). Obwohl der Test sprachliche Anforderungen an die Lernenden stellt, zeigt die Lesefähigkeit, gemessen mit dem Lese- und Verständnis-Test (LGVT), keinen signifikanten Zusammenhang mit der Kompetenz des *wissenschaftlichen Skizzierens* auf (siehe Tab. 55). Deshalb wurden die Variablen der Lesefähigkeit nicht als Prädiktor für eine weitere Regressionsanalyse im Gesamtmodell eingesetzt. Für die Analyse wurde eine einschließende Regression für die Kognition (KFT) und das Strategiewissen zum Experimentieren (NAW) berechnet (siehe Tab. 56).

Tab. 56 Prädiktoren für die Personenfähigkeit

	B	s.e.	Beta (β)	T	p
Regressionskonstante	-1.206	0.0064		-18.844	< 0.001
KFT	0.071	0.008	0.449	9.391	< 0.001
NAW	0.061	0.012	0.253	5.306	< 0.001
R^2	0.381				
f	0.784				

Die Prädiktoren NAW und KFT tragen mit einer Effektstärke von $f^2 = 0.615$ stark zur Varianzaufklärung der Kompetenz des *wissenschaftlichen Skizzierens* bei (siehe Tab. 56) und haben somit nach Cohen (1988) mit > 0.35 eine starke Vorhersagekraft für das *wissenschaftliche Skizzieren*.

Es zeigt sich nach Cohen (1988) ein hochsignifikanter, starker Zusammenhang der Kognition (KFT) ($\beta = 0.449$) und ein hochsignifikanter, moderater Zusammenhang des NAW-Tests ($\beta = 0.253$) mit der Performanz des *wissenschaftlichen Skizzierens* (siehe Tab. 56).

Nach Peterson und Brown (2005) lässt sich das standardisierte Beta (β) mit dem R gleichsetzen, woraus die Effektstärke f der Varianzaufklärung ermittelt werden kann.

Die Kognition zeigt mit $f^2 = 0.253$ eine hohe Effektstärke für die Varianzaufklärung. Der NAW zeigt mit $f^2 = 0.068$ eine schwache Effektstärke für die Varianzaufklärung (Cohen, 1988) (siehe Tab. 56).

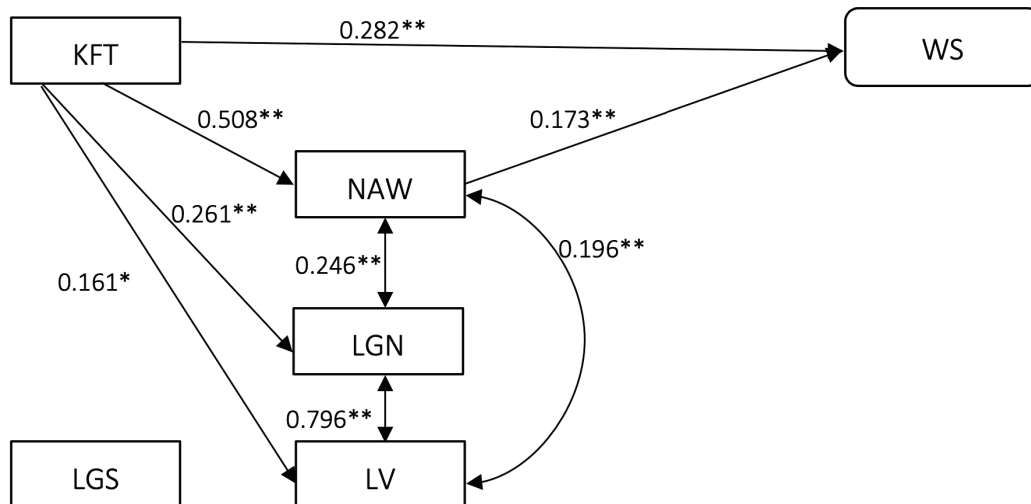


Abb. 17 Strukturgleichungsmodell (Mplus) mit KFT als übergeordneter Faktor als Modell dargestellt, wobei * = $p < 0.1$, ** = $p < 0.001$

Tab. 57 Kennwerte für das Strukturgleichungsmodell (Mplus)

Item Kennwerte	KFT als übergeordneter Faktor (Abb. 17)	allen Faktoren gleich gewichtet (Abb. 18)
AIC	7524.013	9605.510
BIC	7594.456	9683.995
RMSEA (Estimate)	0.000	0.000
CFI	1.000	1.000
TLI	1.000	1.000

Um den Zusammenhang der externen Faktoren mit der Performanz des *wissenschaftlichen Skizzierens* (WS) zu prüfen, wurden zwei Strukturgleichungs-Analysen durchgeführt (siehe Abb. 17 und 18). Untersucht wurde, inwiefern die Kognition (KFT) direkt Einfluss auf die Lesefähigkeit (LGVT) und das Strategiewissen zum Experimentieren (NAW) hat und wie diese untereinander korrelieren. Außerdem wurde der Zusammenhang aller Faktoren mit der Kompetenz des *wissenschaftlichen Skizzierens* geprüft. Bei der Lesefähigkeit (LGVT) wurden die Unterkategorien Leseverständnis (LV), Lesegenauigkeit (LGN) und Lesegeschwindigkeit (LGS) untersucht. Die Testgüte wurde bei beiden Analysen durch die Parameter CFI, TLI und RMSEA bestätigt (siehe Tab. 57)

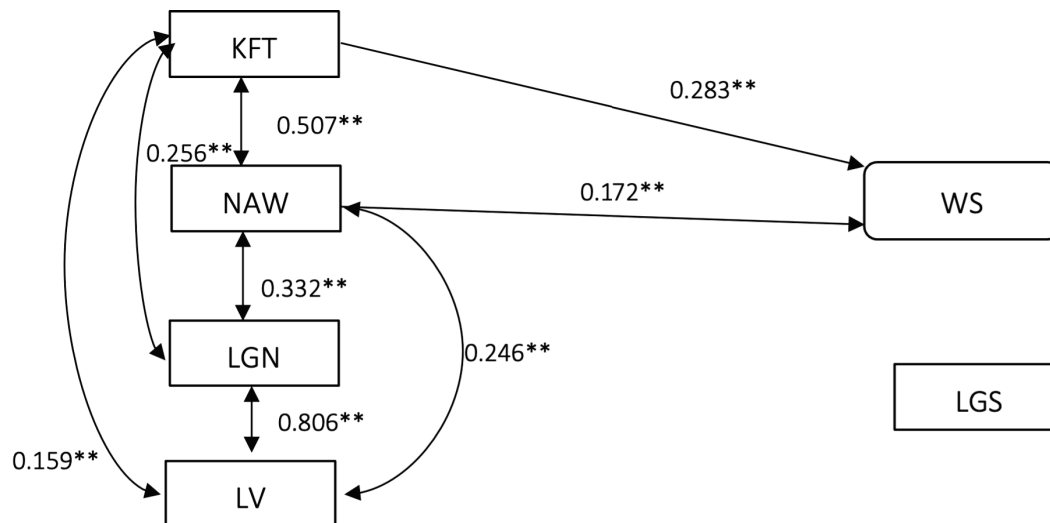


Abb. 18 Strukturgleichungsmodell (Mplus) mit allen Faktoren gleich gewichtet als Modell dargestellt wobei * = $p < 0.1$, ** = $p < 0.001$

Die Untersuchung der Modellpassung zwischen einem Modell, bei dem die Kognition als übergeordneter Einfluss (siehe Abb. 17 und Tab. 57 links) gewichtet wird, und einem Modell, bei dem alle Faktoren gleich gewichtet werden (siehe Abb. 18 und Tab. 57 rechts), wurde mit den Maßen AIC und BIC auf deskriptiver Ebene vorgenommen. Die signifikant kleineren Werte dieser Parameter galten als Entscheidungskriterium für das Modell mit einer übergeordneten Kognition (Geiser, 2011) (siehe Abb. 17).

Da das Modell mit einem übergeordneten Faktor „Kognition“ (siehe Abb. 17) die bessere Modellpassung aufweist (vergleiche Tab. 57 links und rechts) wird auf die ausführliche Beschreibung des anderen Modells (siehe Abb. 18) verzichtet.

Die Lesegeschwindigkeit (LGS) zeigt keinen Zusammenhang mit der Kompetenz des *wissenschaftlichen Skizzierens* und korreliert nicht mit den anderen externen Faktoren. Die Kognition hat ebenfalls keinen Einfluss darauf (siehe Abb. 17). Das bedeutet: Diese Studie ist messinvariant gegenüber der Geschwindigkeit, mit der Schülerinnen und Schüler lesen. Die Kognition, gemessen mit dem KFT, zeigt sowohl einen signifikanten und bedeutenden Zusammenhang mit der Kompetenz des *wissenschaftlichen Skizzierens* (0.282) als auch mit dem Strategiewissen zum Experimentieren (NAW) (0.508) (siehe Abb. 17) (Cohen, 1988). Das Strategiewissen zum Experimentieren, gemessen mit dem NAW, zeigt ebenfalls einen signifikanten, wenn auch weniger bedeutenden Zusammenhang mit der Kompetenz des *wissenschaftlichen Skizzierens* (0.173) (Cohen, 1988). Die Lesegenauigkeit (LGN) zeigt keinen signifikanten Zusammenhang mit der Kompetenz des *wissenschaftlichen Skizzierens*. Die Kognition

weist einen bedeutenden, nicht hochsignifikanten Zusammenhang mit der Lesegenauigkeit (0.261) wie auch mit dem Leseverständnis (LV) (0.161) auf. Zusammenhänge bestehen zudem zwischen der Lesegenauigkeit und dem NAW (0.246), zwischen dem Leseverständnis (LV) und der Lesegenauigkeit (LGN) (0.796) sowie zwischen dem Leseverständnis und dem NAW (0.196). Alle anderen Zusammenhänge sind weder signifikant noch bedeutend (<0.1) (siehe Abb. 17).

Ausgehend von den Auswertungen in diesem Kapitel erfolgt eine interpretative Beantwortung der F4: Das Leseverständnis (LV) und die Lesegenauigkeit (LGS) korrelieren sehr stark (siehe Abb. 17). Das überrascht nicht, denn wer nicht genau liest, versteht auch das Gelesene nicht. Der Zusammenhang der Lesekompetenz mit der Kompetenz des *wissenschaftlichen Skizzierens* ist vernachlässigbar. Der Zusammenhang des NAWs mit der Lesefähigkeit zeigt auf, dass der NAW leicht durch die Lesefähigkeit beeinflusst wird. Die Aufgabentexte sind weitgehend kurz, jedoch relevant für die korrekte Lösung der Aufgaben.

Der wenig bedeutende Zusammenhang des Strategiewissen zum Experimentieren mit der Kompetenz des *wissenschaftlichen Skizzierens* gegenüber dem praktischen Untersuchen in der Studie ExKoNawi (Bonetti et al., 2018) (siehe Kapitel 4.4) zeigt, dass das *wissenschaftliche Skizzieren* zwar mit Strategiewissen zum Experimentieren zusammenhängt, wohl aber ein anderes Konstrukt gemessen wird, wobei das Verständnis der Variablenkontrollstrategie (siehe Kapitel 3.2.2.3) nicht zwangsläufig vorausgesetzt wird. Dies wird dadurch gestützt, dass die Kognition und der NAW einen signifikanten, bedeutenden Zusammenhang zeigen (siehe Kapitel 4.4), wonach der Zusammenhang des NAW mit der Kompetenz des *wissenschaftlichen Skizzierens* teilweise auf die Kognition zurückzuführen ist.

Die Korrelation zwischen NAW und den aus einer Videoanalyse gewonnenen Daten der prozessorientierten Leistungen zeigt, dass der NAW kein geeignetes Ersatzmaß zur Bewertung experimenteller Beobachtungen darstellt (Emden, 2011). Auch scheint der NAW kein Ersatzmaß für die Kompetenz des *wissenschaftlichen Skizzierens* zu sein. Der signifikante und bedeutende Zusammenhang zwischen Kognition und der Kompetenz des *wissenschaftlichen Skizzierens* wird auch durch die Literatur gestützt, wonach ein bedeutender Zusammenhang zwischen Kognition (KFT) und der Performanz bei experimentellen Paper-Pencil-Aufgaben besteht (Wellnitz, 2012), da die Aufgabeninhalte wenig curricular sind und angenommen werden darf, dass mit dem Test

zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* teilweise allgemeine kognitive Fähigkeiten gemessen werden.

Es wurde ein hochsignifikanter und starker Leistungsunterschied der Kompetenz des *wissenschaftlichen Skizzierens* zwischen den Jahrgangsstufen gefunden. Da die Aufgabeninhalte zwar curricular abgestützt sind, aber nicht wesentlich vom Fachwissen abhängen, wird vermutet, dass teilweise allgemeine kognitive Fähigkeiten gemessen werden. Gestützt wird diese Vermutung durch einen hochsignifikanten Zusammenhang (0.449) (siehe Tab. 56) zwischen den kognitiven Fähigkeiten (KFT) und der Kompetenz des *wissenschaftlichen Skizzierens*.

7.5.2 Fach- und Methodenwissen (F5)

Zur Beantwortung der F5 „Besteht ein Zusammenhang zwischen dem relevanten Fach- und Methodenwissen und der Kompetenz des *wissenschaftlichen Skizzierens*?“ wurden die Schülerleistungen im wissenschaftlichen Skizzieren (erreichtes Niveau im jeweiligen Item) mit den Schülerleistungen im Multiple-Select Test über das Fach- und Methodenwissen verglichen, indem drei Kategorien erstellt wurden (siehe Abb. 19).



Abb. 19 Darstellung der Varianteneinteilung der erbrachten Performanz in Fach- und Methodenwissen und dem wissenschaftlichen Skizzieren

Die Fragestellungen bei den Aufgabentypen „Beobachtung“ und „Vergleich“ fokussierten die Beobachtung beziehungsweise den Vergleich von Phänomenen. Die Fragen des Aufgabentyps „Experimentieren“ ermittelten das fachmethodische Wissen zur Variablenkontrolle bei der Untersuchung von Zusammenhängen. Beide korrekten Antworten mussten gegeben werden, damit die Antwort als korrekt galt.

Die Schülerperformanz im *wissenschaftlichen Skizzieren* wurde mit der Schülerperformanz im Fach- und Methodenwissen verglichen, indem drei Varianten erstellt

wurden (siehe Abb. 19). Die erste Variante umfasst die Lernenden, die im *wissenschaftlichen Skizzieren* hohe, jedoch im Fach- und Methodenwissen niedrige Leistungen erbrachten (siehe Abb. 19, dunkelgrau). Die zweite Variante umfasst diejenigen Lernenden, die eine niedrige Performance im *wissenschaftlichen Skizzieren*, jedoch hohe Leistungen im Fach- und Methodenwissen zeigten (siehe Abb. 19, hellgrau). Der Rest umfasst alle anderen Lernenden (siehe Abb. 19, weiß), d. h. diejenigen, die in beiden Kategorien jeweils hohe, tiefe oder mittelmäßige Leistungen erbrachten.

Für die Varianteneinteilung in die Variante 1 wurden drei Untergruppen von Lernenden berücksichtigt, wodurch die Randgruppen berücksichtigt werden konnten. Für die Variante 1 waren dies die Untergruppen:

- i. *wissenschaftliches Skizzieren* Niveau 3 und Fach- und Methodenwissen Niveau 0
- ii. *wissenschaftliches Skizzieren* Niveau 3 und Fach- und Methodenwissen Niveau 1
- iii. *wissenschaftliches Skizzieren* Niveau 2 und Fach- und Methodenwissen Niveau 0

Für die Einteilung in die Variante 2 erfolgte analog zur Einteilung in die Variante 1.

Tab. 58 erbrachte Leistungen im wissenschaftlichen Skizzieren und Multiple-Select-Test zum Fach- und Methodenwissen in Prozent

Item	BR_B in %	EI_B in %	SB_B in %	BR_V in %	EI_V in %	SB_V in %	BR_E in %	EI_Z in %	SB_E in %
Aufgabentyp	Beobachtung			Vergleich			Experimentieren		
wissenschaftliches Skizzieren niedrig, Fachwissen hoch (Variante 1)	23.5	29.3	28.2	19.8	14.8	8.6	8.6	2.4	7.2
Rest	73.5	69.1	71.0	78.6	84.1	89.1	57.8	46.6	62.7
wissenschaftliches Skizzieren hoch, Methodenwissen niedrig (Variante 2)	3.0	1.6	0.8	1.6	1.1	2.3	33.6	51.0	30.1

Tab. 58 zeigt den Prozentsatz der Lernenden je Variante gemäß Abb. 19 je Item. Die markant unterschiedliche Variantengröße beruht auch auf der unterschiedlichen Anzahl gestellter Fragen im Fach- und Methodenwissenstest (siehe Kapitel 6.3.4).

Bei den Aufgabentypen „Beobachtung“ und „Vergleich“ ist der Anteil der Lernenden groß, die zwar in den Kompetenzen des *wissenschaftlichen Skizzierens* tief abschnitten, jedoch hohe Resultate im Multiple-Select Test zeigen. Die kleinere Anzahl Lernender in dieser Variante beim Aufgabentyp „Vergleich“ ist einerseits darauf zurückzuführen, dass bei diesem Aufgabentyp eine Frage mehr bei den Multiple-Select Fragen gestellt wurde, somit das Mittelfeld demensprechend grösser ist, da für die Varianteneinteilung die Randgruppen berücksichtigt wurden. Andererseits deutet es auch darauf hin, dass die Lernenden mit tiefen Kompetenzen im *wissenschaftlichen Skizzieren* im Aufgabentyp „Vergleich“ die Multiple-Select-Fragen schlechter beantwortet haben als im Aufgabentyp „Beobachtung“. Beim Aufgabentyp „Vergleich“ im Kontext „Säuren und Basen“ (SB_V) ist der Anteil der Lernenden, die trotz tiefer Kompetenz im *wissenschaftlichen Skizzieren* die Multiple-Select-Fragen gut geantwortet haben, deutlich geringer als bei den anderen beiden Kontexten dieses Aufgabentyps. Dies kann daher rühren, dass einerseits für diese Aufgabe zwei richtige Antworten angekreuzt werden mussten, und dass andererseits die Multiple-Select-Fragen dieses Kontextes schwieriger waren als die der anderen beiden Kontexte.

Im Aufgabentyp „Experimentieren“ wird mit den Multiple-Select-Fragen das fachmethodische Wissen zur Variablenkontrolle bei der Untersuchung im konkreten Kontext erhoben. Die Lernenden mussten zudem beide korrekten Antworten ankreuzen, um die Aufgabe korrekt zu lösen. Des Weiteren gab es nur zwei Multiple-Select Fragen, wodurch das Mittelfeld als Ganzes kleiner ist. Auffällig ist, dass bei diesem Aufgabentyp zwar kaum Lernende mit tiefen Kompetenzen im *wissenschaftlichen Skizzieren* hohe Resultate im Multiple-Select-Teil erzielten, jedoch viele mit hohen Kompetenzen im *wissenschaftlichen Skizzieren* tiefe Resultate im Multiple-Select Test hatten.

Die unterschiedlichen Zusammenhangsmuster sind auch dem Umstand geschuldet, dass bei den Aufgabentypen „Beobachtung“ und „Vergleich“ mit dem MS-Test etwas anderes gemessen wird als beim Aufgabentyp „Experimentieren“. Bei den Aufgabentypen „Beobachtung“ und „Vergleich“ wurde vor allem das Fachwissen erhoben, wobei beim Aufgabentyp „Experimentieren“ der Fokus auf der Methodenkompetenz lag.

Die unterschiedlichen Zusammenhangsmuster könnten sich daraus erklären, dass Problemstellung und Inhalte der Aufgaben ähnlich bzw. nicht ähnlich waren. Ebenso könnten auch die unterschiedlichen Kodierungen zu den Unterschieden geführt haben.

Zur Bestätigung der Resultate der Gruppierungen (siehe Tab. 58) und, um den Zusammenhang zwischen den Kompetenzen im *wissenschaftlichen Skizzieren* und den Resultaten des Multiple-Select-Tests (Anzahl korrekt gelöster Fragen) zu ermitteln, wurde ein Kendall-Tau-Test (siehe Tab. 59) durchgeführt.

Tab. 59 Kendall-Tau-Test zwischen den Kompetenzen im wissenschaftlichen Skizzieren und den Resultaten des Multiple-Select-Tests zum Fach- und Methodenwissen

	BR_B	EI_B	SB_B	BR_V	EI_V	SB_V	BR_E	EI_Z	SB_E
	Beobachtung			Vergleich			Experimentieren		
Kendall-Tau	0.332	0.612	0.559	0.303	0.316	0.125	-0.412	-0.725	-0.572
Signifikanz	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.37	< 0.001	< 0.001	< 0.001

Zu den Aussagen, die aus den prozentualen Angaben der Vergleiche zwischen der Kompetenz des *wissenschaftlichen Skizzierens* und den Resultaten des Multiple-Select-Tests (siehe Tab. 58) hervorgegangen sind, lassen sich weitere Aussagen aus den Resultaten des Kendall-Tau-Tests (siehe Tab. 59) machen.

Hohe Zusammenhänge zwischen den Aufgaben, wie im Aufgabentyp „Beobachtung“ in den Kontexten „Eis schmelzen“ (EI_B) und „Säuren und Basen“ (SB_B), zeigen an, dass Lernende mit hohen Kompetenzen im *wissenschaftlichen Skizzieren* auch gute Resultate im Multiple-Select-Test erreichen. Die Zusammenhänge im Aufgabentyp „Vergleich“ sind geringer als diejenigen im Aufgabentyp „Beobachtung“. Im Aufgabentyp „Vergleich“ im Kontext „Säuren und Basen“ (SB_V) ist der Zusammenhang nicht signifikant. Im Aufgabentyp „Experimentieren“ ist der Zusammenhang negativ.

Der Zusammenhang zwischen dem *wissenschaftlichen Skizzieren* und dem Methodenwissen des Aufgabentyps „Experimentieren“ ist stark negativ, im Kontext „Eis schmelzen“ (EI_Z) sogar sehr stark negativ (siehe Tab. 59). Dieses Item ist im Bereich des *wissenschaftlichen Skizzierens* das einfachste (siehe Abb. 10 und 11). Die Multiple-Select-Fragen des Aufgabentyps „Experimentieren“ sind schwieriger als die der anderen beiden Aufgabentypen (siehe Kapitel 6.3.4).

Eine mögliche Erklärung wäre die Kopplung der Variablenkontrollstrategie (VKS) an allgemeine kognitive Fähigkeiten, messbar mit dem KFT (Heller & Perleth, 2000). Künsting et al. (2008) konnten eine Korrelation mit Einsatz der VKS und dem KFT wie auch eine Korrelation der VKS und dem Lernerfolg zeigen (siehe Kapitel 3.2.2.3). Schwichow und Nehring (2018) jedoch haben dieselbe Korrelation zwischen dem KFT und der VKS nicht bestätigen können. Lernende mit hohem Vorwissen profitieren jedoch mehr vom Einsatz der VKS als Schüler mit geringem Vorwissen (Künsting et al., 2008) (siehe Kapitel 3.2.2.3).

7.5.3 externe Validierung einzelner Aufgabentypen (F6)

Zur Beantwortung von F6 „Besteht ein Unterschied zwischen den Zusammenhängen der verschiedenen Aufgabentypen des *wissenschaftlichen Skizzierens* mit der Kognition (KFT) und dem Strategiewissen zum Experimentieren (NAW)?“ werden die Zusammenhänge der beiden externen Variablen Kognition (KFT) und Strategiewissen zum Experimentieren (NAW) für die Aufgabentypen des *wissenschaftlichen Skizzierens* für die Aufgabentypen „Beobachtung“ und „Vergleich“ (siehe Tab. 61) und für den Aufgabentyp „Experimentieren“ (siehe Tab. 62) unabhängig voneinander mittels einer zweidimensionalen Rasch-Analyse betrachtet. Die zweidimensionale Rasch-Analyse in Bezug auf Teilskalen ist aufgrund des Forschungsdesigns (jeder Lernende löste nur zwei Aufgaben des Aufgabentyps „Experimentieren“) zwar als kritisch einzuschätzen, dennoch lässt diese Analyse eine erste Abschätzung einer möglichen Zweidimensionalität der Kompetenz des *wissenschaftlichen Skizzierens* zu. Die Rasch-Analyse zeigte für die Berechnungen eine zufriedenstellende Modellpassung auf (siehe Tab. 60 (Wu et al., 2007)).

Tab. 60 Item-Kennwerte der 2D-Rasch-Analyse

N	376
Infit-Werte	$0.94 < \text{infit} < 1.11$
Outfit-Werte	$0.89 < \text{outfit} < 1.11$
T-Wert	$-1.2 < T < 1.4$
EAP/PV Reliabilität	0.692/0.586
Item-Separation-Reliabilität	0.964
Variance	0.708/0.978
Geschätzte Parameter	30
Discrimination	0.54-0.66

Tab. 61 Prädiktoren für die Personenfähigkeit, 2D Betrachtung für die Aufgabentypen „Beobachtung“ und „Vergleich“

	B	s.e.	Beta (β)	T	p
Regressionskonstante	-1.583	0.085		-18.556	< 0.001
KFT	0.070	0.01	0.360	6.951	< 0.001
NAW	0.071	0.015	0.239	4.621	< 0.001
R ²	0.27				

Mit dem Daten der zweidimensionalen Rasch-Analyse wurde die Regressionsanalyse, die Zusammenhänge von NAW und KFT mit den Kompetenzen des *wissenschaftlichen Skizzierens* aufteilt, in den zwei möglichen Dimensionen untersucht.

Die Kognition (KFT) ist grundsätzlich der stärkere Prädiktor als das Strategiewissen zum Experimentieren, gemessen mit dem NAW (Cohen, 1988) (siehe Tab. 56, 61 und 62). Die Kognition zeigt im Aufgabentyp „Experimentieren“ einen hohen Zusammenhang zur Kompetenz *des wissenschaftlichen Skizzierens* (siehe Tab. 62) (Cohen, 1988). Bei den Aufgabentypen „Beobachtung“ und „Vergleich“ (siehe Tab. 61) liegt der Zusammenhang zwischen der Kognition und dem *wissenschaftlichen Skizzieren* nach Cohen (1988) im Übergang von einem moderaten zu einem starken Zusammenhang.

Tab. 62 Prädiktoren für die Personenfähigkeit, 2D Betrachtung für den Aufgabentyp „Experimentieren“

	B	s.e.	Beta (β)	T	p
Regressionskonstante	-0.916	0.095		-9.634	< 0.001
KFT	0.1	0.011	0.443	8.919	< 0.001
NAW	0.071	0.017	0.207	4.169	< 0.001
R ²	0.328				

Auf der Basis der Resultate in der Tab. 56, 61 und 62 kann eine interpretative Beantwortung der F6 gemacht werden: Eine mögliche Interpretation für den leicht höheren Zusammenhang der Kognition mit dem *wissenschaftlichen Skizzieren* im Aufgabentyp „Experimentieren“ gegenüber den anderen beiden Aufgabentypen könnte in Anlehnung an Helmke und Schrader (2009) (siehe Kapitel 4.4) ein Qualitätskriterium der höheren kognitiven Aktivierung beim Aufgabentyp „Experimentieren“ sein.

Obwohl der Aufgabentyp „Experimentieren“ für die Schülerinnen und Schüler leichter zu bewältigen ist (siehe Abb. 10), könnte dieser Aufgabentyp durch die geforderte Darstellung der VKS (siehe Kapitel 3.2.2.3) die Lernenden auf einer kognitiven Ebene mehr herausfordern. Es zeigt sich aber nach Fisher (1921) keine statistische Signifikanz ($p = 0.371$) in den Unterschieden der Zusammenhänge der Kognition mit dem *wissenschaftlichen Skizzieren*, unterschieden nach den Aufgabentypen.

Wie in Kapitel 4.4 aufgeführt, zeigt das Untersuchen im Hands-On-Setting des Projekts ExKoNawi (Bonetti et al., 2018) einen mittleren Zusammenhang mit der Kognition. Der deutlich stärker wirkende Zusammenhang des *wissenschaftlichen Skizzierens* mit der Kognition, insbesondere des Aufgabentyps „Experimentieren“, zeigte auf, dass bei der Darstellung einer Planung einer VKS in einer *wissenschaftlichen Skizze* kognitive Aspekte eine markant zentralere Rolle einnehmen als im Hands-On-Setting.

Das Strategiewissen zum Experimentieren zeigt sich grundsätzlich als schwächerer Prädiktor als die Kognition (Cohen, 1988) (siehe Tab. 56, 61 und 62). Obwohl der Zusammenhang des Strategiewissens zum Experimentieren mit dem *wissenschaftlichen Skizzieren* in den Aufgabentypen „Beobachtung“ und „Vergleich“ (siehe Tab. 62) leicht höher ist als im Aufgabentyp „Experimentieren“, können alle Zusammenhänge nach Cohen (1988) als moderat eingestuft werden. Die Teilkompetenz, *wissenschaftliche Skizzen* der Planung einer Variablenkontrolle wie im Aufgabentyp „Experimentieren“ anzufertigen, zeigt einen geringfügig kleineren Zusammenhang mit dem Strategiewissen zum Experimentieren, gemessen mit dem NAW, als dies die Kompetenzbereiche des *wissenschaftlichen Skizzierens* beim Beobachten und Beobachten von Vergleichen (Aufgabentypen „Beobachtung“ und „Vergleich“) aufweisen.

Die Items des NAW zeigen die Kompetenz der Lernenden auf, die Teilprozesse Idee, Beobachtung und Ergebnis der VKS zu erfassen (siehe Kapitel 3.2.2.3 und 4.4). Der empirisch gemessene geringfügig kleinere Zusammenhang des *wissenschaftlichen Skizzierens* im Aufgabentyp „Experimentieren“ mit dem Strategiewissen zum Experimentieren könnte darauf hinweisen, dass das Verständnis des VKS in diesem Aufgabentyp für die Erstellung der *wissenschaftlichen Skizze* weniger relevant ist als in den beiden anderen Aufgabentypen. Gemäß der Itemkonstruktion sollte eigentlich der Aufgabentyp „Experimentieren“ einen (höheren) Zusammenhang mit dem Strategiewissen zum Experimentieren aufweisen, da dieser darauf abzielt, eine kontrollierte

Experimentieranordnung zu skizzieren. Dagegen spielt diese Kontrolle bei den Aufgabentypen „Beobachtung“ und „Vergleich“ keine Rolle. Dieses Ergebnis entspricht nicht den Erwartungen. Jedoch zeigt sich nach Fisher (1921) keine statistische Signifikanz ($p = 0.761$) in den Unterschieden der Zusammenhänge für das Strategiewissen zum Experimentieren mit dem *wissenschaftlichen Skizzieren*, unterschieden nach den Aufgabentypen.

Der negative Zusammenhang der Kompetenz des *wissenschaftlichen Skizzierens* mit dem Methodenwissen zur VKS im Aufgabentyp „Experimentieren“, gemessen durch den MS-Test, zeigt ebenfalls die Tendenz auf, dass Lernende zwar eine *wissenschaftliche Skizze* der VKS erstellen können, aber die mit dem Multiple-Select-Test gemessene Methodenkompetenz zur VKS nicht vollständig verstanden haben (siehe Kapitel 7.5.2).

8 Zusammenfassende Diskussion und Ausblick

Der erste Teil dieses Kapitels setzt sich in einer Diskussion mit den Resultaten der Studie auseinander, wobei nicht nur die abschließende Beantwortung der Forschungsfragen und die Grenzen der Validierungsaspekte im Zentrum stehen, sondern auch Möglichkeiten zur Verbesserung der Validierung aufgegriffen werden. Außerdem werden die Grenzen der Studie aufgezeigt. Der zweite Teil des Kapitels enthält einen Ausblick, in dem aufgezeigt wird, welche Fragen sich neu ergeben und welche Forschungsschritte mit einem validierten Testinstrument zum *wissenschaftlichen Skizzieren* folgen könnten.

8.1 Zusammenfassende Diskussion

Die Kompetenz, *wissenschaftliche Skizzen* anzufertigen, wurde bislang weder systematisch untersucht, noch liegt ein Messinstrument vor, mit welchem diese Kompetenz erfasst werden kann. Außerdem ist nicht untersucht, inwiefern sich die Kompetenz des *wissenschaftlichen Skizzierens* in verschiedenen Kompetenzbereichen unterscheidet. Das in dieser Arbeit a-priori entwickelte Kompetenzstrukturmodell des *wissenschaftlichen Skizzierens* bildet die Grundlage für das darauf aufbauende Messinstrument zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens*. Die Aufgaben zu den unterschiedlichen Kompetenzbereichen des *wissenschaftlichen Skizzierens* wurden zur Messung der (Teil-)Kompetenzen des *wissenschaftlichen Skizzierens* entwickelt. Die Validierung erfolgte nach Messick (1995) mittels Expertenbefragung, für das Kompetenzstrukturmodell inhaltlich (bezüglich der Qualitätsmerkmale) und für die Testkonstruktion ökologisch. Für die strukturelle Validität wurde die Testgüte überprüft. Aus der Ermittlung der Messinvarianz konnte auf die Generalisierbarkeit geschlossen werden. Die externe Validierung dient der Einordnung der Modellierung und Abgrenzung der Kompetenz des *wissenschaftlichen Skizzierens*.

Zur Konstruktion des Kompetenzstrukturmodells und zur Testkonstruktion und deren bei der Validierung wurden zwei Forschungsfragen formuliert:

F1: Sind die Qualitätsmerkmale des entwickelten Kompetenzstrukturmodells inhaltlich valide?

F2: Sind die für die Testkonstruktion verwendeten Kontexte und Aufgaben ökologisch valide?

Bei der Konstruktion eines Kompetenzstrukturmodells des *wissenschaftlichen Skizzierens* wurde literaturbasiert vorgegangen. Der inhaltlichen Validität der Qualitätsmerkmale des zu erarbeitenden Kompetenzstrukturmodells wurde bereits bei der Entwicklung Rechnung getragen, indem sowohl induktive wie auch deduktive Ansätze berücksichtigt und a-priori entwickelt wurden. Der induktive Ansatz, bei dem aus bereits vorliegenden Schülerskizzen relevante Qualitätsmerkmale für *wissenschaftliche Skizzen* abgeleitet wurden, kam ebenfalls zur Anwendung. Ergänzend zum deduktiven, forschungsbasierten Ansatz wurde eine Expertenbefragung von Fachdidaktikerinnen und Fachdidaktikern der Chemie durchgeführt. Die Expertenbefragung zeigt auf, dass die gewählten Qualitätsmerkmale des Kompetenzstrukturmodells und deren Umsetzung im Aufgabenset relevant für die Kompetenz des *wissenschaftlichen Skizzierens* sind. Die gewählten Kriterien zur Analyse der *wissenschaftlichen Skizzen* (Kategorien und Facetten im Kompetenzstrukturmodell) zeigten aus Sicht der Expertinnen und Experten eine hohe Passgenauigkeit mit der Kompetenz des *wissenschaftlichen Skizzierens*. Die einzelnen Qualitätsmerkmale wurden im Schnitt mit einer Bewertung von „eher ja“ bis zu „unverzichtbar“ bezeichnet. Die Facetten „Eindeutigkeit“, „Idealisierung“ und „fachliche Korrektheit“ wurden von der Expertengruppe im Schnitt als wichtig erachtet. Die Qualitätsmerkmale des entwickelten Kompetenzstrukturmodells sind aus Sicht der Expertinnen und Experten inhaltlich valide.

Nur die Qualitätsmerkmale des Kompetenzstrukturmodells wurden von einer Expertenrunde aus Fachdidaktikerinnen und Fachdidaktiker evaluiert, nicht jedoch die Gruppierung der Qualitätsmerkmale in Facetten. Deren Zusammenstellung erfolgte aufgrund der thematischen Zusammengehörigkeit der einzelnen Qualitätsmerkmale. Die Beschreibungen der Progressionsstufen und die Festlegung, welches Qualitätsmerkmal in welchem Umfang (gar nicht, teilweise, größtenteils) für eine bestimmte Progressionsstufe erfüllt werden muss, wurde ebenfalls nicht einer Expertengruppe vorgelegt, und auch auf ein Expertenrating für die Kodiermanuals der einzelnen Items wurde verzichtet, da diese auf die Facetten im Kompetenzstrukturmodell aufbauen und sehr kontextabhängig sind. Solche zusätzlichen Evaluationen könnten zu einer Verbesserung der inhaltlichen Validierung (siehe Kapitel 6.7.1) führen.

Die Kontextualisierung und die Aufgabentypen (Beobachtung, Vergleich und Experimentieren) stellen die Grundstruktur des neu entwickelten Tests dar. Bei der Kontextualisierung der Testaufgaben wurde die Curricula des Lehrplan 21 (D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz, 2016b) insofern berücksichtigt, dass Kontexte aus dem Anfangsunterricht gewählt wurden. Zur Überprüfung der ökologischen Validität (siehe Kapitel 6.7.2) wurde eine Befragung von Lehrpersonen der Sekundarstufe I sowohl in Bezug auf die Kontexte und Inhalte als auch auf die Aufgabentypen der Testaufgaben durchgeführt. Die im Test berücksichtigten Kontexte fanden bei über der Hälfte der 41 Lehrpersonen „regelmäßig“ Anwendung im Unterricht, bei den anderen wurde „ab und an“ auf die Kontexte zurückgegriffen. Der Aufgabentyp „Beobachtung“ wurde sehr häufig im Unterricht angewandt. Die Aufgabentypen „Vergleich“ und „Experimentieren“ wurden weniger häufig im Unterricht eingesetzt, obwohl auch diese curricular durch den Lehrplan 21 (D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz, 2016b) vorgegeben sind. Die ökologische Validität (siehe Kapitel 6.7.2) des Tests ist aufgrund der Resultate der Lehrpersonenbefragung gegeben. Die für die Testkonstruktion verwendeten Kontexte sind ökologisch valide. Auch bei den Aufgaben des Tests darf die ökologische Validität als gegeben betrachtet werden, auch wenn die Aufgabentypen „Vergleich“ und „Experimentieren“ nur bei einem Drittel der Befragten regelmäßig, bei einem weiteren Drittel ab und zu im Unterricht eingesetzt werden.

Auf eine Befragung von Fachdidaktikerinnen und Fachdidaktikern wurde in Bezug auf die ökologische Validierung (siehe Kapitel 6.7.2) verzichtet; dies, weil vor allem Lehrpersonen darüber Auskunft geben können, welche Themenbereiche sie persönlich im Unterricht konkret behandeln.

In der Arbeit wurden insgesamt fünfzehn Items konstruiert und pilotiert, basierend auf drei unterschiedlichen Aufgabentypen und fünf Kontexten für die Kompetenzmessung. Durch unterschiedliche Kontextualisierungen können sowohl die unterschiedlichen Interessen der Lernenden als auch unterschiedliche Schwierigkeitsgrade der Kontexte berücksichtigt werden. Aufgrund von Überlegungen zur Differenzierung des Antwortformats, zur Kodierbarkeit und zu curricularen Ansprüchen wie auch zur ökologischen Validität wurden schließlich neun Items aus drei Kontexten in der Hauptstudie eingesetzt.

Im Folgenden wird die Testgüte (siehe Kapitel 6.7.3), die Messinvarianz und die Generalisierbarkeit (siehe Kapitel 6.7.4) des neu entwickelten Messinstruments anhand der F3 diskutiert.

F3: Erfüllt das Messinstrument die Testgüte, die Messinvarianz und Generalisierbarkeit?

Alle in der Hauptstudie eingesetzten Items liegen im offenen Antwortformat vor. Die Analyse der Testgüte des neu entwickelten Testinstruments zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* zeigt, dass mit der Bearbeitung von jeweils zwei Items pro Aufgabentyp (total sechs Items wurden gelöst) die Kompetenz des *wissenschaftlichen Skizzierens* mit einer guten Reliabilität (0.709) gemessen werden kann. Die guten Item-Reliabilitäten des durch die Rasch-Modellierung berechneten Modells zeigen, dass Aussagen sehr zuverlässig getroffen werden können.

Die Werte der Reliabilität liegen in einem ähnlichen Bereich wie bei anderen Studien mit offenen Antwortformaten. Eine Verbesserung der Reliabilität und somit der Testgüte (siehe Kapitel 6.7.3) könnte durch die Erhöhung der Itemzahl, durch eine Überarbeitung bzw. Präzisierung der bereits eingesetzten Items wie auch durch eine bessere Schulung der Kodierhelferinnen und -helfer erreicht werden. Die Überarbeitung des Kodiermanuals sowie die mögliche Nutzung weiterer Indikatoren, die in der vorliegenden Studie aufgrund zu niedriger Interrater-Reliabilität ausgeschlossen werden mussten, könnten zu einer weiteren Verbesserung der Reliabilität des gesamten Tests führen. Die Überarbeitung des Kodiermanual sollte mit dem Fokus auf die Präzisierung der Ankerbeispiele und die Formulierung der Kodierung gemacht werden mit dem Ziel, die Interrater-Reliabilität zu erhöhen.

Infit- und Outfitwerte von eins zeigen die Unidimensionalität des Modells auf. Wegen der bereits sehr guten Infit- und Outfitwerte von nahezu eins steht eine Verbesserung derselben bei einer Überarbeitung des Tests und damit in diesem Bereich der Testgüte (siehe Kapitel 6.7.3) nicht im Vordergrund.

Die Personenstreuung ist mit einer Varianz von unter 0.5 relativ klein. Eine Vergrößerung der Varianz könnte in einer Anschluss-Studie durch eine Erhöhung der Anzahl Lernenden sowie durch eine größere Vielfalt in Bezug auf die Leistungsstärke der teilnehmenden Lernenden erreicht werden und so zu einer Erhöhung der Testgüte be-

züglich der Personenstreuung führen. Die Trennung der einzelnen Items ist auch auf der Ebene der Niveaus gut.

Aufgrund der Testwerte erreicht das Messinstrument die Testgüte und erfüllt dadurch diesen Teil der strukturellen Validität (siehe Kapitel 6.7.3).

Für einen Modellvergleich als Bestandteil der strukturellen Validität (siehe Kapitel 6.7.3) bezüglich der Dimension ist die Datenbasis ungenügend, da die Lernenden pro Aufgabentyp nur zwei Aufgaben gelöst haben. Die Auswertungen deuten auf eine mögliche Mehrdimensionalität der Kompetenz des *wissenschaftlichen Skizzierens* hin, dies in Abhängigkeit des Aufgabentyps. Die Aufgabentypen „Beobachtung“ und „Vergleich“ könnten dieselbe Dimension abbilden. Eine Erhöhung der Itemzahl pro Dimension und beispielweise die Wiederaufnahme der nach der Vorpilotierung verworfenen Kontexte könnten konkretere Aussagen über die Modellierung der Kompetenz des *wissenschaftlichen Skizzierens*, wie die mögliche Bestätigung des vermuteten 2D-Modells („Beobachtung“ und „Vergleich“ als eine Dimension, „Experimentieren“ als zweite Dimension) ermöglichen und dadurch die strukturelle Validität (siehe Kapitel 6.7.3) erhöhen.

Im Folgenden werden die Itemschwierigkeiten als Teil der strukturellen Validität auf Itemebene diskutiert. Dabei werden sowohl Aufgabentyp- als auch Kontextabhängigkeiten dargelegt. Wird nur die Kategorie „*Wissenschaftlichkeit der Skizze*“ berücksichtigt, kann der Unterschied der Itemschwierigkeiten zwischen den Aufgabentypen „Beobachtung“ und „Vergleich“ analysiert werden, da beide die gleichen Anforderungen für diese Kategorie im Progressionsmodell aufweisen. Inhaltlich stellt der Aufgabentyp „Vergleich“ im Gegensatz zu anderen empirischen Resultaten höhere Schwierigkeiten als der Aufgabentyp „Beobachtung“. Der Grund dafür ist wahrscheinlich, dass bei einem „Vergleich“ zwei Phänomene erkannt, miteinander verglichen und dargestellt werden müssen, wohingegen beim Aufgabentyp „Beobachtung“ nur ein Phänomen beobachtet und dargestellt werden muss.

Der Aufgabentyp „Experimentieren“ stellt die niedrigsten Anforderungen an die Lernenden. Gründe dafür, dass die Lernenden den Aufgabentyp „Experimentieren“ leichter in einer *wissenschaftlichen Skizze* darstellen, könnten sein, dass sie bei diesem Aufgabentyp keinen Ablauf eines Phänomens darstellen (also keine Bildabfolge machen müssen), sondern das Vorgehen der Variablenkontrollstrategie in einer einzigen Skizze darstellen können.

„Säuren und Basen“ stellt wie erwartet den schwierigsten Kontext dar, da die Lernenden aus dem Alltag weniger damit vertraut sind als mit den Kontexten „Brausetablette“ und „Eis schmelzen“. Eine zusätzliche Anforderung an die Lernenden sind Farbbänderungen, die nur in diesem Kontext relevant sind.

Für die Prüfung der Generalisierbarkeit (siehe Kapitel 6.7.4) des Tests wird die Messinvarianz analysiert: DIF-Analysen bezüglich Personenmerkmalen zeigen, dass das Aufgabenset messinvariant bezüglich des Geschlechtes und der Jahrgangsstufe der Sekundarstufe I ist. Es gibt weder eine Bevorzugung noch eine Benachteiligung bei den Lernenden der 7. oder 9. Jahrgangsstufe und auch nicht bezüglich des Geschlechts. Auch hinsichtlich der Anzahl Sprachen (nur Deutsch, eine oder mehrere Fremdsprachen), welche die Lernenden zuhause sprechen, sind die Items weitgehend messinvariant. Einzig das Item EI_Z (Kontext „Eis schmelzen“, Aufgabentyp „Experimentieren“) deutet auf einen Vorteil für Lernende, die nicht mehrere Fremdsprachen sprechen, hin.

Bei der Messinvarianz (siehe Kapitel 6.7.4) in Bezug auf die Schulstufe der Lernenden (Gymnasium oder Sekundarstufe I) kann davon ausgegangen werden, dass weder eine Bevorzugung noch eine Benachteiligung einer der beiden Gruppen vorliegt, obwohl zwei Items auf eine Bevorzugung bzw. Benachteiligung der Lernenden des Gymnasiums hindeuten. Um aussagekräftige Darlegungen diesbezüglich machen zu könnten, müsste die Stichprobe der Lernenden des Gymnasiums vergrößert werden. Eine Testoptimierung im Bereich der Differenzierung könnte Aufschluss über die Messinvarianz bezüglich des besuchten Schulniveaus geben. Mittels einer Definierung der Stichprobe, wie bspw., dass alle Lernenden aus demselben Kanton stammen, oder mittels genauer Erfassung des besuchten Schulniveaus (Sek A, B, C oder Gymnasium), könnten neben konkreten Aussagen in Bezug auf die Messinvarianz des besuchten Schulniveaus (siehe Kapitel 6.7.4) zusätzlich Aussagen über die Fähigkeiten der Lernenden gemacht werden.

Aufgrund der weitgehenden Messinvarianz des Tests können daraus generalisierbare Schlüsse (siehe Kapitel 5.7.4) gezogen werden: Der Test ist geeignet zur Erhebung der Kompetenz des *wissenschaftlichen Skizzierens* in der Sekundarstufe I.

Zusammenfassend kann gesagt werden: Der entwickelte und validierte Test weist gute Gütekriterien wie eine hohe Reliabilität, gute Infit- und Outfit-Werte sowie eine gute Itemseparierung auch auf der Ebene der Niveaus auf. Der Test ist messinvariant be-

züglich des Geschlechtes, der Jahrgangsstufe, der Sprache sowie der Personenfähigkeiten und ist homogen, was aus den guten Infit- und Outfitwerten hervorgeht. Er erweist sich als fair und ermöglicht dadurch die Generalisierbarkeit. Gründe dafür könnten darin liegen, dass die Anwendung der Video-Vignetten die Resultate des Tests einheitlicher macht, als wenn die Lernenden selbst Experimente durchführen würden.

Die Festlegung der Schwellenwerte erfolgte normativ aufgrund der Progressionen des Kompetenzstrukturmodells, und sie bildet die empirischen Datenlage ab. Die Festlegung wurde nicht durch ein Expertenrating überprüft und stellt daher eine Schwachstelle der strukturellen Validierung (siehe Kapitel 6.7.3) dar. Es darf jedoch davon ausgegangen werden, dass dadurch, dass Kompetenzausprägungen der einzelnen Facette (Zusammenfassung einzelner Qualitätsmerkmale) die empirische Datenlage abbilden, die Festlegung der Schwellenwerte und somit die Einteilung der Niveaus strukturell valide (siehe Kapitel 6.7.3) ist.

In der Folge werden die Frage des Zusammenhangs mit externen Variablen wie auch die externe Validierung (siehe Kapitel 6.7.5) besprochen. Neben der Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* wurden aus den in der Begleitstudie eingesetzten Frage- und Erhebungsbögen auch die Kognition (KFT), die Erfassung des Strategiewissens zum Experimentieren (NAW), die Lesefähigkeit (LGVT) und das Item-abhängige Fach- und Methodenwissen ermittelt. Da bis auf den Test zur Erfassung des Fach- und Methodenwissens die eingesetzten Erhebungsmittel bereits evaluierte Skalen haben, wurde auf eine Modellpassung dieser Skalen verzichtet.

Aus dem Zusammenhang zwischen der Kognition, dem Strategiewissen zum Experimentieren (NAW) und der Lesefähigkeit mit der Kompetenz des *wissenschaftlichen Skizzierens* und damit der externen Validierung (siehe Kapitel 6.7.5) leitet sich die F4 ab:

F4: Besteht ein Zusammenhang zwischen dem *wissenschaftlichen Skizzieren* und der Kognition (KFT), dem Strategiewissen zum Experimentieren (NAW) und der Lesefähigkeit (LGVT)?

Es konnte gemäß den Erwartungen gezeigt werden, dass ein positiver Zusammenhang zwischen der Kompetenz des *wissenschaftlichen Skizzierens* und der Kognition, gemessen mit dem KFT, vorliegt (siehe Hypothese 4.1). Ebenso sind Zusammenhänge

zwischen dem Strategiewissen zum Experimentieren (NAW) und der Kognition und – wenn auch geringere – zwischen dem Strategiewissen zum Experimentieren (NAW) und der Kompetenz des *wissenschaftlichen Skizzierens* vorhanden (siehe Hypothese 4.2). Diese bestätigen die Modellierung und Abgrenzung der Kompetenz des *wissenschaftlichen Skizzierens* und sind Teil der externen Validierung (siehe Kapitel 6.7.5).

Die Prädiktoren NAW und KFT zusammen tragen mit 61.6% stark zur Varianzaufklärung bei. Die Kognition (KFT) allein gemessen hat mit 25.2% eine starke Effektstärke, und das mit einem Paper-and-Pencil gemessene Strategiewissen (NAW) als Disposition macht rund 6.8 % der Varianz des Settings (*wissenschaftliches Skizzieren*) als Performanz aus und zeigt somit eine mittlere Effektstärke.

Forschungsergebnisse aus anderen Studien stützen die hier gemessenen Zusammenhänge (siehe Kapitel 4.4 und 3.2.2.3). Der Zusammenhang der Kognition mit dem NAW mit der naturwissenschaftlichen Aufgabensequenz (praktisches Handeln) und mit der Erkenntnisgewinnung bzw. dem naturwissenschaftlichen Problemlösen zeigt ebenfalls in der Tendenz, dass höhere kognitive Fähigkeiten mit besseren Leistungen beim naturwissenschaftlichen Problemlösen einhergehen. Somit kann vermutet werden, dass *wissenschaftliches Skizzieren* ähnlich von der Kognition abhängt wie das naturwissenschaftlichen Problemlösen.

Die Strukturgleichungsanalyse dieser Studie zeigt ähnliche Sachverhalte der externen Validierung (siehe Kapitel 6.7.5) auf: Die Kognition wirkt wie erwartet signifikant und bedeutend auf das Strategiewissen zum Experimentieren und ebenso signifikant, aber deutlich schwächer, auf die Lesegenauigkeit und auf das Leseverständnis. Die letzten beiden zeigen erwartungsgemäß eine sehr hohe Abhängigkeit voneinander.

Die Kognition wirkt signifikant stark auf die Kompetenz des *wissenschaftlichen Skizzierens* und ebenfalls signifikant, wenn auch weniger stark, auf die naturwissenschaftlichen Arbeitsweisen, gemessen mit dem NAW. Der signifikante, bedeutende, jedoch schwächere Zusammenhang des *wissenschaftlichen Skizzierens* mit dem NAW als mit dem KFT macht deutlich, dass das *wissenschaftliche Skizzieren* zwar Zusammenhänge zum Konstrukt der naturwissenschaftlichen Arbeitsweise hat, aber nicht vollständig damit erklärt werden kann. Der Zusammenhang zwischen NAW und *wissenschaftlichem Skizzieren* hängt deutlich von der Kognition ab. Fachwissen scheint kaum eine Rolle zu spielen. Dies stützen die Forschungsergebnisse von Klos et al., die keine Abhängigkeit vom Fachwissen zum NAW zeigen konnten, und auch die-

jenigen von Schwichow und Nehring (siehe Kapitel 3.1.2.3), welche nur von einer indirekten kausalen Abhängigkeit zwischen Fachwissen und Erkenntnisgewinnung, bei der die Kognition die ausschlaggebende Komponente zu sein scheint, ausgehen.

Die Hypothese 4.1 kann bestätigt werden: Die Kognition steht wie erwartet in einem starken, positiven Zusammenhang mit *wissenschaftlichem Skizzieren* bei allen drei Aufgabentypen. Auch die Hypothese 4.2 kann vollumfänglich bestätigt werden: Das Strategiewissen zum Experimentieren steht ebenfalls wie erwartet in einem mittleren positiven Zusammenhang mit *wissenschaftlichem Skizzieren* bei allen drei Aufgabentypen, wobei das Strategiewissen stark von der Kognition abhängt, was Analogien zu den Studien von Nehring et al. (2020) und Wirth et al. (2005) zulässt (siehe Kapitel 3.1.2.3 und 4.4).

Die Lesefähigkeit zeigt keinen Zusammenhang mit der Kompetenz des *wissenschaftlichen Skizzierens* (siehe H 4.3). Forschungsergebnisse aus anderen Studien stützen diese Feststellung (siehe Kapitel 4.4 und 3.2.2.3). Es wurde kein sprachlicher Einfluss auf das Strategiewissen beim Experimentieren (NAW) – gemessen mit der verbalen Skala des KFT – nachgewiesen. Ein Zusammenhang zwischen der Lesegeschwindigkeit und der Kompetenz des *wissenschaftlichen Skizzierens* konnte nicht bestätigt werden, was den Schluss zulässt, dass das Messinstrument invariant bezüglich Lesegeschwindigkeit (LGS) der Lernenden ist. Die Strukturgleichungsanalyse ergab außerdem keinen signifikanten Zusammenhang zwischen dem Leseverständnis (LV) und der Lesegenauigkeit (LGN) mit der Kompetenz des *wissenschaftlichen Skizzierens*. Sie trägt dadurch zur externen Validierung bei (siehe Kapitel 6.7.5). Die Lesefähigkeit zeigt erwartungsgemäß keinen Zusammenhang mit der Kompetenz des *wissenschaftlichen Skizzierens*, wodurch auch die Hypothese 4.3 bestätigt werden kann.

Die Einflüsse des Kontexts auf die Leistung der Lernenden wurde nicht vertieft untersucht. Auch die Frage, ob der Kontext je nach Aufgabentyp einen anderen Einfluss auf die Leistungen der Lernenden hat, stand nicht im Zentrum der Untersuchungen. Aufgrund ähnlicher Niveauschwellen für die gleichen Aufgabentypen kann aber davon ausgegangen werden, dass der Kontext keinen expliziten Einfluss hat, sondern vor allem Kodierungsunterschiede zwischen den einzelnen Kontexten (wie z.B. unterschiedliche Anzahl Indikatoren) Einfluss auf die Festlegung des Leistungsniveaus der Lernenden haben. Es wurde bei der Aufgabenentwicklung zwar explizit darauf geachtet, dass die Items möglichst wenig von Vor- und Fachwissen abhängen, jedoch

wurde der Einfluss von Vor- und Fachwissen des Kontextwissens auf die Leistung im *wissenschaftlichen Skizzieren* – und somit dieser Teil der externen Validierung (siehe Kapitel 6.7.5) – nicht ermittelt.

Das Fach- und Methodenwissen wurde als weiterer externer Faktor zur externen Validierung (siehe Kapitel 6.7.5) untersucht. Neben der mit offenem Antwortformat durch den validen Test erfassten Kompetenz des *wissenschaftlichen Skizzierens* wurde mit einem geschlossenen Antwortformat das Fach- und Methodenwissen erhoben, und zwar spezifisch zu jedem Item kontext- und aufgabentypabhängig. Bei den Aufgabentypen „Beobachtung“ und „Vergleich“ beziehen sich die Fragen auf Beobachtungs- und Vergleichsaspekte. Beim Aufgabentyp „Experimentieren“ stützen sich die Fragen auf das Methodenwissen der Variablenkontrollstrategie.

In Bezug auf einen Zusammenhang des Fach- und Methodenwissens mit der *Kompetenz des wissenschaftlichen Skizzierens* und der externen Validierung (siehe Kapitel 6.7.5) leitet sich die F5 ab:

F5: Besteht ein Zusammenhang zwischen dem relevanten Fach- und Methodenwissen und der Kompetenz des *wissenschaftlichen Skizzierens*?

Bei den Aufgabentypen „Beobachtung“ und „Vergleich“ zeigte das relevante Fach- und Methodenwissen, der Erwartung entsprechend, einen mittleren, positiven Zusammenhang mit der Kompetenz des *wissenschaftlichen Skizzierens* (siehe Hypothese 5.1). Rund ein Viertel der Lernenden zeigte trotz tiefer Performanz im *wissenschaftlichen Skizzieren* hohe Leistungen im relevanten Fach- und Methodenwissen. Es gab kaum Lernende, die im Fach- und Methodenwissen-Test gut abschnitten, aber tiefe Leistungen im *wissenschaftlichen Skizzieren* zeigten. Die Lernenden verfügen vermutlich über Beobachtungskompetenzen (Methodenkompetenz), um die relevanten Vorgänge der Reaktion oder die relevanten Details des Phänomens oder der Phänomene zu erfassen, wie auch über das nötige kontextspezifische Fachwissen, das mit den Multiple-Select-Fragen überprüft wird, und zeigen daher im Multiple-Select-Test eine gute Leistung. Die Lernenden sind aber teilweise nicht in der Lage, ihre Beobachtungen in *wissenschaftlichen Skizzen* darzustellen.

Für den Aufgabentyp „Experimentieren“ zeigen, wie vermutet, Multiple-Select-Fragen einen negativen Zusammenhang zur Kompetenz des *wissenschaftlichen Skizzierens* (siehe Hypothese 5.2). Diese Fragen greifen das Methodenwissen der Variablen-

kontrollstrategie (VKS) auf, von welcher Sodian und Meyer zeigten, dass Kinder zwar über ein implizites Wissen darüber verfügen, jedoch ihre Entscheidung nicht auf der Grundlage der VKS begründen können (siehe Kapitel 3.2.2.3). Knapp die Hälfte der Lernenden zeigte zwar eine hohe Performanz im *wissenschaftlichen Skizzieren* von der Planung zur Untersuchung von Zusammenhängen, verfügte jedoch nicht über das notwendige Methodenwissen der Variablenkontrolle (VKS). Auffällig ist, dass vor allem die Lernenden der 7. Jahrgangsstufe nicht über dieses Methodenwissen verfügten. Wie Schwichow und Nehring (2020) zeigen, geht eine höhere Kompetenzausprägung in der VKS mit höherem Fachwissen und höherer Jahrgangsstufe einher (siehe Kapitel 3.1.2.3), wobei das höhere Fachwissen stark an höhere Kognition gekoppelt zu sein scheint. Die Altersabhängigkeit der VKS kann durch die Resultate von Sodian und Mayer erklärt werden. Dass Lernende im Gymnasium, bedingt durch die Unterrichtsform, oftmals über das größere Fachwissen verfügen, wäre – gestützt auf die Analyse von Schwichow und Nehring – eine Erklärung für die besseren Leistungen sowohl im Fach- und Methodenwissen als auch in der der Kategorie „*Wissenschaftlichkeit der Skizze*“ für den Aufgabentyp „Experimentieren“, da sowohl das Fach- und Methodenwissen als auch die Kategorie „*Wissenschaftlichkeit der Skizze*“ stark von der VKS abhängig sind.

Zur Erweiterung der externen Validität (siehe Kapitel 6.7.5) wäre zu untersuchen, ob die Lernenden in der Lage wären, die Variablenkontrollstrategie anhand erstellter wissenschaftlicher Skizzen durchzuführen. Dies könnte auch ohne, dass die Lernenden über die Methodenkompetenz der Variablenkontrolle verfügen, möglich sein.

Bei der Kompetenz des *wissenschaftlichen Skizzierens* im Aufgabentyp „Experimentieren“ ist die Performanz der Lernenden höher als bei den anderen Aufgabentypen. Für die Darstellung einer Planung einer VKS (Aufgabentyp „Experimentieren“) reicht, anders als bei den anderen beiden Aufgabentypen, ein einziges Bild anstelle einer Bildabfolge. Dies könnte ein Grund dafür sein, dass diese Form der *wissenschaftlichen Skizze* Lernenden grundsätzlich leichter fällt als die Darstellung von Beobachtungen, sei es von einem einzelnen oder von mehreren Phänomenen. Die Fragen zum Methodenwissen (Multiple-Select-Fragen) der Variablenkontrollstrategie (Aufgabentyp „Experimentieren“) stellen für die Lernenden die größere Herausforderung dar als die Fach- und Methodenfragen zu den beiden anderen Aufgabentypen, da diese

inhaltlich-methodisch wie auch wegen der Ratewahrscheinlichkeit (zwei richtige und vier falsche Antworten) anspruchsvoller sind.

Die Hypothese 5.1 kann bestätigt werden: Das relevante Fach- und Methodenwissen der Aufgabentypen „Beobachtung“ und „Vergleich“ steht in einem mittleren positiven Zusammenhang mit *wissenschaftlichem Skizzieren* der Aufgabentypen „Beobachtung“ und „Vergleich“.

Die Hypothese 5.2 kann nicht vollständig bestätigt werden: Das relevante Fach- und Methodenwissen des Aufgabentyps „Experimentieren“ steht in einem negativen Zusammenhang mit *wissenschaftlichem Skizzieren des* Aufgabentyps „Experimentieren“. Grund dafür könnte sein, dass vor allem bei jüngeren Lernenden die Methodenkompetenz der Planung eines Experiments noch nicht immer ausgeprägt vorhanden war (siehe Kapitel 3.2.2.3).

Um noch aussagekräftigere Einschätzungen bezüglich des Zusammenhangs zwischen Fach- und Methodenwissen und der Kompetenz des *wissenschaftlichen Skizzierens* abzugeben und dadurch die externe Validität (siehe Kapitel 6.7.5) zu erhöhen, wäre zu untersuchen, inwiefern Lernende die Multiple-Select-Fragen zu den Aufgabentypen ohne die dazugehörige Video-Vignette bewerkstelligen könnten. Es wurde nicht untersucht, inwiefern Fragen zur Methodenkompetenz des Beobachtens und Vergleichen für Lernende anspruchsvoller wären, wie dies bei den Fragen zur Methodenkompetenz der VKS der Fall ist.

Ausgehend von der Möglichkeit einer Zweidimensionalität der Kompetenz des *wissenschaftlichen Skizzierens* und dadurch der strukturellen Validität (siehe Kapitel 6.7.3) wurde trotz nicht vollständig passendem Studiendesign der Zusammenhang von Strategiewissen zum Experimentieren, gemessen mit dem NAW, und von Kognition, gemessen mit dem KFT, mit der Kompetenz des *wissenschaftlichen Skizzierens* in Abhängigkeit des Aufgabentyps untersucht. Dabei wurden, gemäß der möglichen Zweidimensionalität, die Aufgabentypen „Beobachtung“ und „Vergleich“ zusammen als eine Dimension und der Aufgabentyp „Experimentieren“ als zweite Dimension betrachtet.

In Bezug auf eine mögliche Unterscheidung des *wissenschaftlichen Skizzierens* in Teilskalen lässt sich die F6 ableiten:

F6: Besteht ein Unterschied zwischen den Zusammenhängen der verschiedenen Aufgabentypen des *wissenschaftlichen Skizzierens* mit der Kognition (KFT) und dem Strategiewissen zum Experimentieren (NAW)?

Die Kognition (KFT) ist grundsätzlich der stärkere Prädiktor als das Strategiewissen zum Experimentieren, gemessen mit dem NAW. Dies gilt sowohl für die Betrachtung aller Aufgabentypen zusammen als auch für die Betrachtung in der Unterscheidung von Kompetenzbereichen des *wissenschaftlichen Skizzierens*. Im Aufgabentyp „Experimentieren“ ist der Zusammenhang zwischen der Kognition und dem *wissenschaftlichen Skizzieren* leicht höher als bei den anderen beiden Aufgabentypen. Die Effektstärke des Zusammenhangs nach Cohen (siehe Kapitel 7.5.3) ist für die Aufgabentypen „Beobachtung“ und „Vergleich“ im Übergang vom moderaten zum starken Zusammenhang. Über alle Aufgabentypen hinweg und bei der einzelnen Betrachtung des Aufgabentyps „Experimentieren“ ist die Effektstärke für den Zusammenhang der Kognition mit dem *wissenschaftlichen Skizzieren* groß (siehe Hypothese 6.1). Bei diesem Aufgabentyp könnte es zu einer höheren kognitiven Aktivierung führen, was den höheren Zusammenhang mit der Kognition begründen könnte (siehe Kapitel 4.4 und 6.5.3) (Helmke & Schrader, 2009).

Die Unterschiede der Effektstärke des Zusammenhangs der Kognition mit dem *wissenschaftlichen Skizzieren* erwiesen sich jedoch als nicht statistisch signifikant. Eine größere Stichprobe könnte über die mögliche Zweidimensionalität, und dadurch über die strukturelle Validität (dieses Kapitel 6.7.3), Aufschluss geben.

Das Strategiewissen zum Experimentieren ist grundsätzlich der schwächere Prädiktor als die Kognition. Dies gilt sowohl für die Betrachtung aller Aufgabentypen zusammen als auch für die Betrachtung in der Unterscheidung von Kompetenzbereichen des *wissenschaftlichen Skizzierens*. Die Effektstärke des Zusammenhangs nach Cohen (siehe Kapitel 7.5.3) ist für alle Aufgabentypen zusammen wie auch einzeln betrachtet im moderaten Bereich. Die Kompetenzbereiche des *wissenschaftlichen Skizzierens* der Aufgabentypen „Beobachtung“ und „Vergleich“ zeigten, entgegen den Erwartungen, einen leicht stärkeren Zusammenhang mit dem Strategiewissen zum Experimentieren, gemessen mit dem NAW, als der Kompetenzbereich, *wissenschaftliche Skizzen* einer Planung einer Variablenkontrolle wie im Aufgabentyp „Experimentieren“ anzufertigen (siehe Hypothese 6.1). Erkennen und Vergleichen von Phänomenen scheint leicht stärker vom Strategiewissen zum Experimentieren abzuhängen als von der Ko-

gnition. Der geringere Zusammenhang der Kognition (gemessen mit der KFT) bei diesen Teilprozessen könnte darauf zurückgeführt werden, dass vor allem Beobachtungskompetenzen und die Darstellung von Beobachtetem im Fokus stehen und die Lernenden weniger auf kognitive Prozesse zurückgreifen müssen.

Die Unterschiede der Effektstärke des Zusammenhangs des Strategiewissen zum Experimentieren mit dem *wissenschaftlichen Skizzieren* erwiesen sich jedoch als nicht signifikant (siehe Kapitel 6.5.3). Eine größere Stichprobe könnte darüber und über die strukturelle Validität (siehe Kapitel 5.7.3) Aufschluss geben.

Die Hypothese 6.1 kann nur teilweise bestätigt werden: Die Kompetenz des *wissenschaftlichen Skizzierens* lässt sich bezüglich der beiden untersuchten Aufgabentypen (Kompetenzbereiche) in Bezug auf das Fach- und Methodenwissen unterscheiden (siehe F5 und Kapitel 6.5.2). Der Zusammenhang von Kognition (KFT) und Strategiewissen zum Experimentieren (NAW) mit der Kompetenz des *wissenschaftlichen Skizzierens* unterscheidet sich bei den Aufgabentypen „Beobachtung“ und „Vergleich“ im Vergleich zum Aufgabentyp „Experimentieren“ zwar, dieser Unterschied erwies sich jedoch nicht als signifikant, wodurch die Hypothese 6.1 in Bezug auf die Kognition und das Strategiewissen zum Experimentieren (NAW) nicht genügend aussagenkräftig ist. Ob ein zweidimensionales Modell bei der Kompetenz des *wissenschaftlichen Skizzierens* vermutet wird, kann jedoch aufgrund der zu geringen Datenmenge des Forschungsdesigns nicht abschließend beantwortet werden. Eine Änderung des Forschungsdesigns, durch das die Anzahl der Items für die möglichen zwei Dimensionen erhöht würde, könnte eine mögliche Bestätigung des vermuteten 2D-Modells liefern und dadurch die strukturelle Validität (siehe Kapitel 6.7.3) erhöhen.

Zusammenfassend konnte bezüglich der externen Validität (siehe Kapitel 6.7.5) gezeigt werden, dass es sich bei der Kompetenz des *wissenschaftlichen Skizzierens* um ein Konstrukt handelt, das wie erwartet von der Kognition sowie – wenn auch weniger – vom Strategiewissen zum Experimentieren abhängt. Die Kompetenz kann nicht ausschließlich durch Kognition und Strategiewissen zum Experimentieren abgebildet werden. Das *wissenschaftliche Skizzieren* lässt sich nicht als Kompetenzbereich der experimentellen Kompetenz erfassen, da *wissenschaftliches Skizzieren* in jedem Teilprozess des Experimentierprozesses erforderlich und zudem transferfähig ist. Die Kompetenz des *wissenschaftlichen Skizzierens* ist ein neues Konstrukt, das sich weder durch die Kognition noch durch das Strategiewissen zum Experimentieren, das Fach-

und Methodenwissen, die experimentelle Kompetenz oder die Repräsentationskompetenz vollständig abbilden oder voraussagen lässt.

Im Folgenden werden Aspekte für mögliche Folgestudien besprochen, die zu einer Verbesserung der unterschiedlichen Validierungsaspekte (siehe Kapitel 6.7) führen können: Die inhaltliche (siehe Kapitel 6.7.1), die ökologische (siehe Kapitel 6.7.2) und die externe (siehe Kapitel 6.7.5) Validierung des Testinstruments wie auch die Validierungsaspekte Testgüte (siehe Kapitel 6.7.3), Messinvarianz und Generalisierbarkeit (siehe Kapitel 6.7.4) könnten in unterschiedlichen Bereichen erweitert und dadurch verbessert werden. Die inhaltliche Validierung der Qualitätsmerkmale des entwickelten Kompetenzstrukturmodells sowie die Validierung der Schwellenwerte und dadurch der Festlegung der Niveaus im Kompetenzstrukturmodell könnten nach der Weiterentwicklung erneut durch eine Expertenrunde von Fachdidaktikerinnen und Fachdidaktiker erfolgen. Der Einsatz einer Expertenrunde für die Zuordnung der Qualitätsmerkmale pro Facette im Kompetenzstrukturmodell wie auch für die Beschreibung der Niveaus im Kompetenzstrukturmodell würde zu einer Verbesserung der inhaltlichen Validierung des Kompetenzmodells führen. Die Festlegung der Schwellenwerte könnte ebenfalls durch ein Expertenrating erfolgen und nicht, wie im vorliegenden Fall, durch einen Vergleich der normativen und der empirischen Datenlage, was die inhaltliche und strukturelle Validität ebenfalls verbessern könnte. Die Ermessensspielräume wie auch die Einflüsse der einzelnen Facetten auf die Zuordnung des Niveaus würden durch eine Validierung der Festlegung der Schwellenwerte verringert werden. Dabei wäre es auch möglich, für eine zusätzliche inhaltliche Validierung die Kodiermanuals der neun Items durch eine Expertenrunde einschätzen zu lassen.

Die ökologische Validierung (siehe Kapitel 6.7.2) der für die Testkonstruktion verwendeten Kontexte und Aufgaben könnte durch eine homogenere Expertenrunde verfeinert werden. Weitere Faktoren, wie Arbeitsort (oder -land), Arbeitserfahrung oder Schulhauskultur, könnten in der Expertenrunde berücksichtigt werden. Außerdem könnten für die ökologische Validierung auch Fachdidaktikerinnen und Fachdidaktiker neben den Lehrpersonen befragt werden, wodurch die ökologische Validierung breiter abgestützt würde.

Die Erfüllung der Validierungsaspekte Testgüte (siehe Kapitel 6.7.3), Messinvarianz und Generalisierbarkeit (siehe Kapitel 6.7.4) könnte durch eine Vergrößerung der Stichprobe verbessert werden. Vor allem die Messinvarianz bezüglich Leistungsni-

veaus war bei vereinzelt Items wegen der geringen Stichprobe an Gymnasiastinnen und Gymnasiasten nicht genügend aussagekräftig. Durch eine größere Stichprobe an Gymnasiastinnen und Gymnasiasten könnten abschließende Aussagen über die Messinvarianz der Leistungsniveaus gemacht und dadurch die Generalisierbarkeit (siehe Kapitel 6.7.4) erhöht werden.

Das Forschungsprojekt lässt auch Fragen in Bezug auf die Generalisierbarkeit (siehe Kapitel 6.7.4) und die externe Validierung offen: Der Einfluss des Kontexts (Brause-tablette, Eis schmelzen, Säuren und Basen) auf die Leistung der Lernenden im *wissenschaftlichen Skizzieren* wurde nur am Rande untersucht. Der Frage, inwiefern das Vorwissen bezüglich eines Kontexts die Leistungen der Lernenden beeinflusst, könnte zum einen durch die Wiederaufnahme der ausgeschiedenen Kontexte konkreter nachgegangen werden oder zum anderen auch durch eine vertiefte Auseinandersetzung mit den drei Kontexten der Hauptstudie. Eine Generalisierbarkeitsstudie (G-Study) könnte Auskunft über die Zuverlässigkeit und Reproduzierbarkeit der Messungen geben und zur Erhöhung der Generalisierbarkeit (siehe Kapitel 6.7.4) beitragen.

Eine weitere Folgestudie könnte die Erweiterung des Tests auf andere Fachkontexte der Chemie zum Inhalt haben, was die Generalisierbarkeit (siehe Kapitel 6.7.4) ebenfalls erhöhen würde. Relevant für die Auswahl von Fachkontexten ist die Unabhängigkeit von Vor- und Fachwissen wie auch die Umsetzbarkeit der Fachkontexte in allen drei Aufgabentypen. Die ökologische Validierung neuer Kontexte sollte berücksichtigt werden. Bereits die in der Vorpilotierung ausgeschiedenen Kontexte bspw. berücksichtigen diese Kriterien.

In einer Folgestudie könnte außerdem ein Fachwissentest im Themenbereich der Kontexte des Messinstruments oder der Curricula des Anfangsunterrichts zur Stützung des Zusammenhangs zwischen dem relevanten Fach- und Methodenwissen und der Kompetenz des *wissenschaftlichen Skizzierens* und dadurch der externen Validierung (siehe Kapitel 6.7.5) beitragen. Die Methodenkompetenz des Beobachtens könnte aufgrund des Settings (beobachten, um zu skizzieren) einen relevanten Einfluss auf die Kompetenz des *wissenschaftlichen Skizzierens* haben. In einer Folgestudie sollte das Erhebungssetting um einen Test zur Erfassung der Beobachtungskompetenz ergänzt werden, um so den Einfluss dieser Kompetenz auf das *wissenschaftliche Skizzieren* einschätzen zu können.

Die Untersuchung der Entwicklung der Variablenkontrollstrategie (VKS) mittels *wissenschaftlicher Skizzen* wäre auch eine mögliche Weiterführung der Studie. Sind die Lernenden in der Lage, eine VKS experimentell umzusetzen, auch wenn sie nicht über die entsprechende Methodenkompetenz verfügen, wie diese Arbeit das hier aufzeigen konnte? Wäre gar die Förderung der Entwicklung der VKS durch das *wissenschaftliche Skizzieren* möglich? Eine Nachfolgestudie mit Fokus auf diese Fragestellungen könnte sehr aufschlussreich sein.

Interessant wäre es, der Modellierung der Kompetenzbereiche des *wissenschaftlichen Skizzierens*, wie der vermuteten Zweidimensionalität („Beobachtung“ und „Vergleich“ als eine Dimension, „Experimentieren“ als zweite Dimension) (siehe auch F6) nachzugehen. Eine Überarbeitung des Forschungsdesigns, bspw. durch eine Erhöhung der Itemzahl für die Dimension des Aufgabentyps „Experimentieren“, wäre dazu erforderlich. Das könnte durch die Wiederaufnahme, der nach der Vorpilotierung verworfenen Kontexte, oder durch eine Neukonzeptionierung des Aufgabentyps in den vorhandenen Kontexten erreicht und dadurch die strukturelle Validität (siehe Kapitel 6.7.3) verbessert werden. Auch eine generelle Erhöhung der Stichprobe könnte verbesserte Erkenntnisse über die nicht statistisch signifikanten Unterschiede in Bezug auf die externen Faktoren Kognition und Strategiewissen zum Experimentieren bringen und dadurch verbesserten Aufschluss über die strukturelle (siehe Kapitel 6.7.3) und externe Validität geben.

In der Arbeit wurden zwei Zellen der matrixartigen Verknüpfung der *naturwissenschaftlichen Arbeitsweisen* und des *wissenschaftlichen Denkens* (siehe Tab. 2) berücksichtigt. Die Frage, inwiefern Aussagen über die Kompetenz des *wissenschaftlichen Skizzierens* auf andere Zellen dieser Matrix übertragen werden können, gilt es noch zu klären. Gerade eine Erweiterung der Kompetenz des *wissenschaftlichen Skizzierens* auf andere Teilprozesse des Experimentierprozesses, die in dieser Studie nicht berücksichtigt wurden (z.B. Auswertung) wäre denkbar, wodurch die Generalisierbarkeit (siehe Kapitel 6.7.4) der Kompetenz des *wissenschaftlichen Skizzierens* erhöht werden könnte. Hier müsste auch die Frage geklärt werden, was noch als *wissenschaftliches Skizzieren*, wie es hier definiert und angewandt wird, angesehen werden kann und welche Skizzen diese relativ engen Grenzen der Definition der Kompetenz des *wissenschaftlichen Skizzierens* überschreiten. Des Weiteren könnte auch die Untersuchung der Transferfähigkeit der Kompetenz des *wissenschaftlichen Skizzierens*

auf Bereiche außerhalb der experimentellen Kompetenzen, wie dem Fachwissen, Fragestellungen für eine Folgestudie bieten und dadurch die externe Validierung (siehe Kapitel 6.7.5) verbreitern.

Eine weitere offene Frage des Forschungsprojekts ist die Übertragbarkeit der Kompetenz des *wissenschaftlichen Skizzierens* auf andere Fächer, wie bspw. die Biologie und die Physik. Hierbei müsste berücksichtigt werden, dass in der Biologie zwischen den Methoden des Betrachtens und des Beobachtens deutlich unterschieden wird. Im vorliegenden Forschungsprojekt wurde die Methode des Betrachtens nicht berücksichtigt, da die Chemie sich vor allem mit Vorgängen der (Stoff)veränderung beschäftigt. Bei einer möglichen Übertragbarkeit auf die Physik stellen sich Fragen nach Abhängigkeiten von Fachwissen oder auch nach der Rolle curricular früh vermittelter Modelldarstellungen (z.B. Schaltschemen) für die Darstellung des Stromkreislaufs.

8.2 Ausblick

Es ist in diesem Forschungsvorhaben gelungen, ein valides Testinstrument zu entwickeln, das die Kompetenz des *wissenschaftlichen Skizzierens* für gewisse Zellen der matrixartigen Verknüpfung von *naturwissenschaftlichen Arbeitsweisen* und *wissenschaftlichem Denken* erfassen kann und somit für weiterführende Erhebungen und Studien zur Verfügung steht – wie etwa für eine Untersuchung der Kompetenz des *wissenschaftlichen Skizzierens* bspw. Stichprobe, die hinsichtlich des unterrichteten Niveaus, der Region oder der Lehrperson definiert ist. Dadurch könnte weiterer Aufschluss über die Fähigkeiten der Lernenden im *wissenschaftlichen Skizzieren* gewonnen werden. Ebenfalls können kausale Zusammenhänge von Personenvariablen der Lernenden untersucht werden. Es kann gemessen werden, was die Lernenden können, wann (zeitlich in der Schullaufbahn) sie ihre Kompetenzen im *wissenschaftlichen Skizzieren* erweitern und auch, welche Unterschiede zwischen Lernenden aus verschiedenen Regionen, Jahrgangs- oder Niveaustufen bestehen.

Auf der Ebene der fachdidaktischen Forschung ist die Kompetenz des *wissenschaftlichen Skizzierens* erstmals definiert und gegenüber anderen Konstrukten abgegrenzt worden. Außerdem liegen erstmals Resultate bezüglich Zusammenhängen des *wissenschaftlichen Skizzierens* zur Kognition, zum Strategiewissen beim Experimentieren und zur Lesekompetenz vor. Die Grenzen der Studie ergeben sich unter anderem im Bereich der möglichen Übertragbarkeit der Resultate auf andere Zellen der Matrix

naturwissenschaftliche Arbeitsweise und *wissenschaftliches Denken*. Dadurch ergeben sich neue Fragen, die in einem neuen Forschungsvorhaben realisiert werden könnten: Zeigt die Kompetenz des *wissenschaftlichen Skizzierens* je nach Zelle der Matrix *naturwissenschaftliche Arbeitsweise* und *wissenschaftliches Denken* andere Schwerpunkte und somit auch andere Zusammenhangsmuster zu externen Variablen, wie der Kognition oder dem Strategiewissen beim Experimentieren, wie es die Resultate dieser Studie vermuten lassen?

Dank des Vorliegens eines validen Messinstruments zur Erfassung der Kompetenz des *wissenschaftlichen Skizzierens* kann in Zukunft in einem neuen Forschungsprojekt in Interventionsstudien die effektive Wirkung von Fördermaßnahmen auf die Kompetenz des *wissenschaftlichen Skizzierens* untersucht werden. Fragen wie: „Welche Maßnahme hat bei welcher Jahrgangsstufe die größte Wirkung?“ können nachgegangen werden.

Auch ein Einsatz im Unterricht in der Zielstufe wird durch das neu entwickelte Progressionsmodell möglich: Das Messinstrument könnte Einsatz in der Volksschule finden, indem Lehrpersonen es zur formativen wie auch summativen Bewertung von *wissenschaftlichen Skizzen* von Lernenden nutzen. Denkbar wäre, für die Volksschule, ausgehend vom Progressionsmodell, ein Bewertungsraster zu entwickeln, mit dem Skizzen in Bezug auf das *wissenschaftliche Skizzieren* zur formativen und summativen Bewertung genutzt werden können. Das *wissenschaftliche Skizzieren* könnte bspw. aber auch zur Erfassung der experimentellen Kompetenz genutzt werden, indem *wissenschaftliche Skizzen* bewusst auf die Entwicklung der experimentellen Fähigkeiten der Lernenden hin (z.B. für die Planung) eingesetzt würden.

Die Ergebnisse dieser Arbeit bieten Ansätze für die Umgestaltung des naturwissenschaftlichen Unterrichts im Bereich des nicht sprachlich-verbalen Dokumentierens. Das Progressionsmodell zur Kompetenz des *wissenschaftlichen Skizzierens* liefert Aussagen über die Schwierigkeit der unterschiedlichen Kompetenzbereiche, abgebildet durch die Aufgabentypen des *wissenschaftlichen Skizzierens*.

Es bietet sich ebenfalls an, in einem weiteren Forschungsprojekt Fragen nach dem Stellenwert des *wissenschaftlichen Skizzierens* nachzugehen, insbesondere auch der Frage, was es bedeutet, wenn in einem anderen Fach das Lernziel lautet, die verbal-sprachlichen Kompetenzen zu fördern.

Zusammenfassend konnte mit diesem Forschungsprojekt ein valides Messinstrument zur Erfassung des *wissenschaftlichen Skizzierens* geschaffen werden, das einerseits in weiteren Forschungsvorhaben noch tiefergreifender und auch erweitert erfasst werden kann, das aber andererseits gerade in der Volkshule bereits zum Einsatz kommen kann und Aufschluss über die Kompetenzen der Lernenden im Bereich des *wissenschaftlichen Skizzierens* bietet.

9 Literaturverzeichnis

- Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 33(2–3), 131–152. [https://doi.org/10.1016/S0360-1315\(99\)00029-9](https://doi.org/10.1016/S0360-1315(99)00029-9)
- Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science*, 333(6046), 1096–1097.
- Alesandrini, K. L. (1981). Pictorial–verbal and analytic–holistic learning strategies in science learning. *Journal of Educational Psychology*, 73(3), 358.
- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. *Educational Psychologist*, 48(2), 87–113. <https://doi.org/10.1080/00461520.2013.775712>
- Arnold, J., Kremer, K., & Mayer, J. (2016). Concept Cartoons als diskursiv-reflexive Szenarien zur Aktivierung des Methodenwissens beim Forschenden Lernen. *Zeitschrift für Didaktik der Biologie (ZDB) - Biologie Lehren und Lernen*, 20, 33–43. <https://doi.org/10.4119/zdb-1636>
- Arnold, J., Wellnitz, N., & Mayer, J. (2010). Beschreibung und Messung von Beobachtungskompetenz bei Schülerinnen und Schülern der Sekundarstufe I. D. Krüger, A. Upmeyer zu Belzen, & S. Nitz, *Erkenntnisweg Biologiedidaktik*, 9, 7–22.
- Azevedo, F. S. (2000). Designing representations of terrain: A study in meta-representational competence. *The Journal of Mathematical Behavior*, 19(4), 443–480.
- Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In PISA 2000 (S. 15–68). Springer.
- Bayrak, C., Hoffmann, L., & Ralle, B. (2015). Sprachliches und fachliches Lernen im Experimentalunterricht. *MNU Ausgabe*, 3(2015), 177–182.
- Bayrak, C., & Ralle, B. (2016). Versuchsprotokolle und Vermittlung von Textsortenkompetenz. In C. Maurer (Hrsg.), *Authentizität und Lernen-das Fach in der Fachdidaktik*. Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Berlin 2015. Universität Regensburg.
- Bergeler, E. (2009). Lernen durch eigenständiges Schreiben von sachbezogenen Texten im Physikunterricht [PhD Thesis]. Technische Universität Dresden.
- Bonetti, A., Gut, C., Metzger, S., & Walpuski, M. (2018). Performanz beim Experimentieren mit und ohne Experimentiermaterial. In C. Mauer (Hrsg.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe*. GDGP Jahrestagung 2017 (Bd. 39, S. 73–76). Universität Regensburg.
- Bonetti, A., Metzger, S., & Gut, C. (2017). Validierung des ExKoNawi-Modells (Experimentelle Kompetenzen in den Naturwissenschaften). In C. Maurer (Hrsg.), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis*. GDGP Jahrestagung 2016 Zürich (Bd. 37, S. 336–339). Universität Regensburg.
- Bortz, J., & Döring, N. (2007). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler: Limitierte Sonderausgabe*. Springer.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Springer.
- Bowen, G. M., & Roth, W.-M. (2002). Why students may not learn to interpret scientific inscriptions. *Research in Science Education*, 32(3), 303–327.
- Brandenburger, M., Mikelskis-Seifert, S., Schwichow, M., & Wilbers, J. (2020). Variablenkontrollstrategien in der Grundschule. In S. Habig (Hrsg.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen*. GDGP Jahrestagung 2019 (Bd. 40, S. 130–133). Uni Duisburg-Essen.

<http://nbn-resolving.de/urn:nbn:de:0111-pedocs-204454>

Brezmann, S. (2004). Erkenntnistätigkeiten im Unterricht-Teil 5: Das Definieren und Vergleichen. *Praxis der Naturwissenschaften Biologie in der Schule*, 28–29.

Britton, L. A., & Wandersee, J. H. (1997). Cutting up Text to Make Moveable, Magnetic Diagrams: A Way of Teaching & Assessing Biological Processes. *The American Biology Teacher*, 59(5), 288–291. <https://doi.org/10.2307/4450310>

Brown, J. H., & Shavelson, R. J. (1996). *Assessing Hands-On Science: A Teacher's Guide to Performance Assessment*. Corwin Press, Inc.

Bryce, T. G. K. (1984). *Techniques for the Assessment of Practical Skills in Foundation Science. Teacher's Guide*. Heinemann.

Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion*. Pearson Deutschland GmbH.

Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. Pearson Deutschland GmbH.

Bybee, R. W. (2002). Scientific Literacy—Mythos oder Realität? *Scientific literacy: der Beitrag der Naturwissenschaften zur allgemeinen Bildung*, 21–43.

Cardella, M. E., Atman, C. J., & Adams, R. S. (2006). Mapping between design activities and external representations for engineering student designers. *Design Studies*, 27(1), 5–24. <https://doi.org/10.1016/j.destud.2005.05.001>

Cardiff. (2008). *TeleForm [Computer software] [Software]*. Lüneburg: Electric paper.

Causey, A. (2016). *Drawn to see: Drawing as an ethnographic method*. University of Toronto Press.

Cavelti, M., Bühner, W., & Berset, T. (2018). Documentation in chemistry teaching—A theory based report of practice; Der Dokumentier-Führerschein—Theoriegestützter Praxisbericht. *Progress in Science Education (PriSE)*, 1(1). <https://doi.org/10.25321/prise.2018.760>

Chang, H.-Y. (2018). Students' representational competence with drawing technology across two domains of science. *Science Education*, 102(5), 1129–1149. <https://doi.org/10.1002/sce.21457>

Chang, H.-Y., & Tzeng, S.-F. (2018). Investigating Taiwanese Students' Visualization Competence of Matter at the Particulate Level. *International Journal of Science and Mathematics Education*, 16(7), 1207–1226. <https://doi.org/10.1007/s10763-017-9834-2>

Chen, Z., & Klahr, D. (1999). All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy. *Child Development*, 70(5), 1098–1120. <https://doi.org/10.1111/1467-8624.00081>

Cheng, M. M. (2018). Students' visualisation of chemical reactions – insights into the particle model and the atomic model. *Chemistry Education Research and Practice*, 19(1), 227–239. <https://doi.org/10.1039/C6RP00235H>

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37–46.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Academic Press.

Cooper, M. M., Stieff, M., & DeSutter, D. (2017). Sketching the invisible to predict the visible: From drawing to modeling in chemistry. *Topics in cognitive science*, 9(4), 902–920. <https://doi.org/10.1111/tops.12285>. 2017

- Cooper, M. M., Underwood, S. M., Hilley, C. Z., & Klymkowsky, M. W. (2012). Development and assessment of a molecular structure and properties learning progression. *Journal of Chemical Education*, 89(11), 1351–1357.
- Cooper, M. M., Williams, L. C., & Underwood, S. M. (2015). Student Understanding of Intermolecular Forces: A Multimodal Study. *Journal of Chemical Education*, 92(8), 1288–1298. <https://doi.org/10.1021/acs.jchemed.5b00169>
- Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, 9(4), 343–363. [https://doi.org/10.1016/S0959-4752\(98\)00051-6](https://doi.org/10.1016/S0959-4752(98)00051-6)
- Crocker, L. (1997). Assessing content representativeness of performance assessment exercises. *Applied Measurement in Education*, 10(1), 83–95.
- Crosland, M. P. (1962). *Historical Studies in the Language of Chemistry*. Courier Corporation.
- D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz. (2016a). Lehrplan 21—Gesamtausgabe—Von der D-EDK Plenarversammlung am 31.10.2014 zur Einführung in den Kantonen freigegebene Vorlage. Bereinigte Fassung vom 29.02.2016. <http://www.lehrplan.ch>
- D-EDK Deutsch-Schweizer Erziehungsdirektoren-Konferenz. (2016b). Lehrplan 21—Natur und Technik mit Physik, Chemie und Biologie—Kompetenzaufbau 3. Zyklus—Von der D-EDK Plenarversammlung am 31.10.2014 zur Einführung in den Kantonen freigegebene Vorlage. Bereinigte Fassung vom 29.02.2016. <http://www.lehrplan.ch>
- Dempsey, B. C., & Betz, B. J. (2001). Biological drawing: A scientific tool for learning. *The American Biology Teacher*, 63(4), 271–279.
- diSessa, A. A. (2002). Students' Criteria for Representational Adequacy. In K. Gravemeijer, R. Lehrer, B. Van Oers, & L. Verschaffel (Hrsg.), *Symbolizing, Modeling and Tool Use in Mathematics Education* (S. 105–129). Springer Netherlands. https://doi.org/10.1007/978-94-017-3194-2_7
- diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and instruction*, 22(3), 293–331.
- diSessa, A. A., & Sherin, B. L. (2000). Meta-representation: An introduction. *The Journal of Mathematical Behavior*, 19(4), 385–398. [https://doi.org/10.1016/S0732-3123\(01\)00051-7](https://doi.org/10.1016/S0732-3123(01)00051-7)
- Dorsch, F., Wirtz, M. A., & Strohmer, J. (2013). *Lexikon der Psychologie* (16. Auflage). Verlag Hans Huber.
- EDK Erziehungsdirektoren-Konferenz (Hrsg.). (2011). *Grundkompetenzen für die Naturwissenschaften—Nationale Bildungsstandards* (freigegeben von der EDK-Plenarversammlung am 16. Juni 2011). https://edudoc.ch/record/96787/files/grundkomp_nawi_d.pdf
- Emden, M. (2011). *Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens: Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I* (Bd. 118). Logos Verlag Berlin.
- Emden, M., & Sumfleth, E. (2012). Prozessorientierte Leistungsbewertung des experimentellen Arbeitens – Zur Eignung einer Protokollmethode zur Bewertung von Experimentierprozessen. *Der mathematische und naturwissenschaftliche Unterricht: MNU*, 65(2), 68–75.
- Engl, L., & Risch, B. (2015). Auswirkung der Protokollierfähigkeit auf die Lernwirksamkeit. In C. Maurer (Hrsg.), *Authentizität und Lernen—Das Fach in der Fachdidaktik. GDCP Jahrestagung 2015 Berlin* (S. 265–267).
- Engl, L., Schumacher, S., Sitter, K., Größler, M., Niehaus, E., Rasch, R., Roth, J., & Risch, B. (2015). Entwicklung eines Messinstrumentes zur Erfassung der Protokollierfähigkeit – initiiert durch Video-Items. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 223–229. <https://doi.org/10.1007/s40573-014-0023-3>

Enyedy, N. (2005). Inventing mapping: Creating cultural forms to solve collective problems. *Cognition and Instruction*, 23(4), 427–466.

Ferber, N. (2014). Entwicklung und Validierung eines Testinstruments zur Erfassung von Kompetenzentwicklung im Fach Chemie in der Sekundarstufe I. Logos Verlag Berlin GmbH.

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. SAGE publications.

Finson, K. D. (2002). Drawing a Scientist: What We Do and Do Not Know After Fifty Years of Drawings. *School Science and Mathematics*, 102(7), 335–345. <https://doi.org/10.1111/j.1949-8594.2002.tb18217.x>

Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, 1, 1–32.

Fleischer, J., Koeppen, K., Kenk, M., Klieme, E., & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. *Zeitschrift für Erziehungswissenschaft*, 16(1), 5–22.

Fleiss, J. L., & Cohen, J. (1973). The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement*, 33(3), 613–619. <https://doi.org/10.1177/001316447303300309>

Frischknecht-Tobler, U., & Labudde, P. (2010). Beobachten und Experimentieren. In P. Labudde (Hrsg.), *Fachdidaktik Naturwissenschaft 1.-9. Schuljahr* (S. 133–148). Haupt.

Gebre, E. H., & Polman, J. L. (2016). Developing young adults' representational competence through infographic-based science news reporting. *International Journal of Science Education*, 38(18), 2667–2687.

Geiser, C. (2011). *Datenanalyse mit Mplus*. Springer.

Giere, R. N., & Moffatt, B. (2003). Distributed cognition: Where the cognitive and the social merge. *Social studies of science*, 33(2), 301–310.

Gilbert, J. K. (2008). Visualization: An Emergent Field of Practice and Enquiry in Science Education. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Hrsg.), *Visualization: Theory and Practice in Science Education* (S. 3–24). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5267-5_1

Goodwin, C. (1994). Professional vision. *American anthropologist*, 96(3), 606–633.

Gott, R., & Welford, G. (1987). The Assessment of Observation in Science. *School science review*, 69(247), 217–227.

Greeno, J. G., & Hall, R. P. (1997). Practicing representation: Learning with and about representational forms. *Phi Delta Kappan*, 78, 361–367.

Gropengießer, H., Kattmann, U., & Krüger, D. (2010). *Biologiedidaktik in Übersichten*. Aulis-Verl.

Groß, K., & Reiners, C. S. (2012). Experimente alternativ dokumentieren. *CHEMKON*, 19(1), 13–20. <https://doi.org/10.1002/ckon.201110165>

Gut, C. (2012). Modellierung und Messung experimenteller Kompetenz: Analyse eines large-scale Experimentiertests (Bd. 134). Logos Verlag Berlin.

Gut, C., Hild, P., Metzger, S., & Tardent, J. (2017). Vorvalidierung des ExKoNawi-Modells. In C. Maurer (Hrsg.), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis. GDGP Jahrestagung 2016 Zürich* (S. 328–331). Universität Regensburg.

Gut, C., & Mayer, J. (2018). Experimentelle Kompetenz. In *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 121–140). Springer.

Gut, C., Metzger, S., Hild, P., & Tardent, J. (2014). Problemtypenbasierte Modellierung und Messung

experimenteller Kompetenzen. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung*, 0(0). <http://www.phydid.de/index.php/phydid-b/article/view/532>

Gwet, K. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical methods for inter-rater reliability assessment*, 1(6), 1–6.

Gwet, K. (2008). Variance estimation of nominal-scale inter-rater reliability with random selection of raters. *Psychometrika*, 73(3), 407–430.

Hammann, M. (2002). *Kriteriengeleitetes Vergleichen im Biologieunterricht*. Studien-Verlag.

Hammann, M. (2004). Kompetenzentwicklungsmodelle. *Der mathematische und naturwissenschaftliche Unterricht*, 57(4), 196–203.

Harle, M., & Towns, M. H. (2013). Students' understanding of primary and secondary protein structure: Drawing secondary protein structure reveals student understanding better than simple recognition of structures. *Biochemistry and Molecular Biology Education*, 41(6), 369–376. <https://doi.org/10.1002/bmb.20719>

Harmon, M. (1997). Performance assessment in IEA's third international mathematics and science study (TIMSS) (International Association for the Evaluation of Educational Achievement, Hrsg.). TIMSS International Study Center.

Harms, U., Mayer, J., Hammann, M., Bayrhuber, H., & Kattmann, U. (2004). Kerncurriculum und Standards für den Biologieunterricht in der gymnasialen Oberstufe. In H.-E. Tenorth (Hrsg.), *Kerncurriculum Oberstufe II. Biologie, Chemie, Physik, Geschichte, Politik* (S. 22–84). Weinheim: Beltz.

Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Heidelberg: Springer Medizin Verlag Heidelberg. https://doi.org/10.1007/3-540-33020-8_9

Heller, K. A., & Perleth, C. (2000). KFT 4-12+ R kognitiver Fähigkeitstest für 4. Bis 12. Klassen, Revision. Beltz Test.

Helmke, A., & Schrader, F.-W. (2009). Qualitätsmerkmale „guten Unterrichts“. *Handbuch der Erziehungswissenschaft*, 701–712. https://doi.org/10.30965/9783657764969_068

Hild, P., Gut, C., & Brückmann, M. (2019). Validating performance assessments: Measures that may help to evaluate students' expertise in 'doing science'. *Research in Science & Technological Education*, 37(4), 419–445. <https://doi.org/10.1080/02635143.2018.1552851>

Hild, P., Kölbach, E., & Metzger, S. (2015). Beobachten lernen—Aufgaben zur Förderung der Beobachtungskompetenz. *Naturwissenschaften im Unterricht. Chemie*, 149.

Hill, M., & Sharma, M. D. (2015). Students' Representational Fluency at University: A Cross-Sectional Measure of How Multiple Representations are Used by Physics Students Using the Representational Fluency Survey. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(6), 1633–1655. <https://doi.org/10.12973/eurasia.2015.1427a>

Hinze, S. R., Rapp, D. N., Williamson, V. M., Shultz, M. J., Deslongchamps, G., & Williamson, K. C. (2013). Beyond ball-and-stick: Students' processing of novel STEM visualizations. *Learning and Instruction*, 26, 12–21. <https://doi.org/10.1016/j.learninstruc.2012.12.002>

Höner, K., & Wenzel, D. (2018). Messung naturwissenschaftlich-experimenteller Problemlösefähigkeit und deren Zusammenhang mit kognitiven und nicht-kognitiven Persönlichkeitsmerkmalen von Schülerinnen und Schülern. ZFdn eingereicht.

Honomichl, K., Risler, H., & Rupprecht, R. (2013). *Wissenschaftliches Zeichnen in der Biologie und verwandten Disziplinen*. Springer.

Hope, G. (2008). The types of drawings that young children produce in response to design tasks. *Design and Technology Education: An International Journal*, 10(1), 43–53.

Hussy, W. (1984). *Denkpsychologie: Ein Lehrbuch*. Informationen Deutsch als Fremdsprache, 14(5–6), 482–483.

Ingold, T. (2011). *Being Alive: Essays on Movement, Knowledge and Description*. Taylor & Francis.

Jee, B. D., Gentner, D., David H. Uttal, Sageman, B., Forbus, K., Manduca, C. A., Ormand, C. J., Shipley, T. F., & Tikoff, B. (2014). Drawing on experience: How domain knowledge is reflected in sketches of scientific structures and processes. *Research in Science Education*, 44(6), 859–883.

Jenßen, L., Dunekacke, S., & Blömeke, S. (2015). Qualitätssicherung in der Kompetenzforschung. Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden* (S. 11–31). Beltz Juventa.

Johnstone, A. H. (1982). *Macro and microchemistry*. Royal soc chemistry, Thomas Graham House.

Jonassen, D. (2003). Using cognitive tools to represent problems. *Journal of research on Technology in Education*, 35(3), 362–381.

Kelley, T. R., & Sung, E. (2017). Sketching by design: Teaching sketching to young learners. *International Journal of Technology and Design Education*, 27(3), 363–386.

Kelly, R. M., Barrera, J. H., & Mohamed, S. C. (2010). An Analysis of Undergraduate General Chemistry Students' Misconceptions of the Submicroscopic Level of Precipitation Reactions. *Journal of Chemical Education*, 87(1), 113–118. <https://doi.org/10.1021/ed800011a>

Kiyokawa, S., Kura, Y., Uesaka, Y., & Manalo, E. (2012). Does Construction of Diagrams Deepen Understanding by Raising Awareness of Insufficiency in Learning? Staging knowledge and experience: how to take advantage of representational technologies in education and training?, 100–102.

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes* (S. xvi, 239). The MIT Press.

Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? In *Pädagogik* (Bd. 56, S. 10–13).

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., & Tenorth, H.-E. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. BMBF.

Klieme, E., Funke, J., Leutner, D., Reimann, P., & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47(2), 179–200.

Klos, S., Henke, C., Kieren, C., Walpuski, M., & Sumfleth, E. (2008). Naturwissenschaftliches Experimentieren und chemisches Fachwissen—zwei verschiedene Kompetenzen. *Zeitschrift für Pädagogik*, 54(3), 304–321.

Kluwe, R. H. (2006). *Informationsaufnahme und Informationsverarbeitung*. Hogrefe.

Kobow, I. (2015). *Entwicklung und Validierung eines Testinstrumentes zur Erfassung der Kommunikationskompetenz im Fach Chemie*. Logos Verlag Berlin GmbH.

Koenen, J. (2014). *Entwicklung und Evaluation von experimentunterstützten Lösungsbeispielen zur Förderung naturwissenschaftlich-experimenteller Arbeitsweisen* (Bd. 171). Logos Verlag Berlin GmbH.

Kohlhauf, L. (2012). *Spielend biologisch beobachten: Messung und Förderung biologischer Beobachtungskompetenz im Vorschulalter*. Logos Verlag Berlin.

- Kohlhauf, L., Rutke, U., & Neuhaus, B. (2010). Beobachten als wissenschaftliche Erkenntnis methode. Entwicklung eines Kompetenzmodells für Kinder ab dem Vorschulalter. *Erkenntnisweg Biologiedidaktik*, 9, 165–178.
- Koller, I., Alexandrowicz, R., & Hatzinger, R. (2012). Das Rasch Modell in der Praxis: Eine Einführung in eRm. UTB.
- Köller, O., Fischer, H., Mayer, J., Sumfleth, E., Hartmann, S., Katzenbach, M., Kauertz, A., Notarp, H., Ropohl, M., & Walpuski, M. (2008). Evaluation der Standards in den Fächern Biologie, Chemie und Physik für die Sekundarstufe I (ESNaS). Band 5: Aufgabenbeispiele für die Aufgabenentwicklung in den Fächern Biologie, Chemie und Physik für den Kompetenzbereich „Erkenntnisgewinnung.
- Koslowski, B. (1996). *Theory and Evidence: The Development of Scientific Reasoning*. MIT Press.
- Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction*, 13(2), 205–226. [https://doi.org/10.1016/S0959-4752\(02\)00021-X](https://doi.org/10.1016/S0959-4752(02)00021-X)
- Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *The Journal of the Learning Sciences*, 9(2), 105–143.
- Kozma, R., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of research in science teaching*, 34(9), 949–968.
- Kozma, R., & Russell, J. (2005). Students becoming chemists: Developing representationl competence. In *Visualization in science education* (S. 121–145). Springer.
- Kraus, M. E., & Stehlik, S. (2008). Protokolle schreiben. *Unterricht Physik*, 19(104), 17–22.
- Kremer, K., Fischer, H. E., Kauertz, A., Mayer, J., Sumfleth, E., & Walpuski, M. (2012a). Assessment of standard-based learning outcomes in science education: Perspectives from the German project ESNAS. *Making it tangible: learning outcomes in science education*, 201–218.
- Kremer, K., Fischer, H. E., Kauertz, A., Mayer, J., Sumfleth, E., & Walpuski, M. (2012b). Assessment of standard-based learning outcomes in science education: Perspectives from the German project ESNAS. *Making it tangible: learning outcomes in science education*, 201–218.
- Kultusministerkonferenz, K. M. K. (2005c). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. Neuwied: Luchterhand.
- Kultusministerkonferenz, K. M. K. (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss*. Neuwied: Luchterhand.
- Kultusministerkonferenz, K. M. K. (2005a). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss*. Neuwied: Luchterhand. <http://www.kmk.org/schul/home.htm>.
- Künsting, J., Thillmann, H., Wirth, J., Fischer, H. E., & Leutner, D. (2008). Strategisches Experimentieren im naturwissenschaftlichen Unterricht. *Psychologie in Erziehung und Unterricht*, 55(1), 1–15.
- Lehrer, R., & Schauble, L. (2006). *Cultivating model-based reasoning in science education*. Cambridge University Press.
- Leisen, J. (2003). Vorgänge und Experimente beschreiben. *Unterricht Physik*, 14(75), 18–21.
- Leuders, T., & Sodian, B. (2013). Inwiefern sind Kompetenzmodelle dazu geeignet kognitive Prozesse von Lernenden zu beschreiben? *Zeitschrift für Erziehungswissenschaft*, 16(1), 27–33.
- Lunetta, V. N., & Novick, S. (1982). *Inquiring and problem-solving in the physical sciences: A sourcebook*. Kendall/Hunt Publishing Company.

- Madden, S. P., Jones, L. L., & Rahm, J. (2011). The role of multiple representations in the understanding of ideal gas problems. *Chemistry Education Research and Practice*, 12(3), 283–293.
- Mäkelä, M., Nimkulrat, N., & Heikkinen, T. (2014). Drawing as a research tool: Making and understanding in art and design practice. *Studies in Material Thinking*, 10, 1–13.
- Mannel, S. (2011). *Assessing Scientific Inquiry: Development and Evaluation of a Test for the Low-performing Stage* (Bd. 111). Logos-Verlag.
- Mannel, S., Walpuski, M., & Sumfleth, E. (2015). Erkenntnisgewinnung: Schülerkompetenzen zu Beginn der Jahrgangsstufe 5 im naturwissenschaftlichen Anfangsunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 99–110. <https://doi.org/10.1007/s40573-015-0028-6>
- Martin, M. O., Mullis, I. V., Foy, P., & Stanco, G. M. (2012). TIMSS 2011 International Results in Science. ERIC.
- Martschinke, S. (1996). Der Aufbau mentaler Modelle durch bildliche Darstellungen: Eine experimentelle Studie über die Bedeutung der Merkmalsdimensionen Elaboriertheit und Strukturiertheit im Sachunterricht der Grundschule. *Zeitschrift Für Pädagogik*, 42(2), 215–232.
- Masters, G. N., & Wright, B. D. (1997). The Partial Credit Model. In W. J. van der Linden & R. K. Hambleton (Hrsg.), *Handbook of Modern Item Response Theory* (S. 101–121). Springer. https://doi.org/10.1007/978-1-4757-2691-6_6
- Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung* (S. 177–186). Springer.
- Mayer, J., & Wellnitz, N. (2014). Die Entwicklung von Kompetenzstrukturmodellen. In *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 19–29). Springer.
- Medina, R., & Suthers, D. (2013). Inscriptions Becoming Representations in Representational Practices. *Journal of the Learning Sciences*, 22(1), 33–69. <https://doi.org/10.1080/10508406.2012.737390>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741–749.
- Metzger, S., & Gut, C. (2017). Symposium: Experimentelle Kompetenzen in den Naturwissenschaften (ExKoNawi). In C. Maurer (Hrsg.), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis. GDGP Jahrestagung 2016 Zürich* (S. 324–327). Universität Regensburg.
- Miller, D. M., & Linn, R. L. (2000). Validation of Performance-Based Assessments. *Applied Psychological Measurement*, 24(4), 367–378. <https://doi.org/10.1177/01466210022031813>
- Moosbrugger, H., & Kelava, A. (Hrsg.). (2012). *Testtheorie und Fragebogenkonstruktion: Mit 66 Abbildungen und 41 Tab. (2., aktualisierte und überarbeitete Auflage)*. Springer.
- Moosbrugger, H., & Kelava, A. (2020). Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien“). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 13–38). Springer. https://doi.org/10.1007/978-3-662-61532-4_2
- Nehring, A. (2014). *Wissenschaftliche Denk- und Arbeitsweisen im Fach Chemie: Eine kompetenzorientierte Modell- und Testentwicklung für den Bereich der Erkenntnisgewinnung*. Logos Verlag Berlin.
- Nehring, A., & Schwichow, M. (2020). Was ist Wissen, was ist Können? Deutungen des Kompetenzbegriffs und deren psychometrische Konsequenzen im Kontext von Fachwissen und Variablenkontrollstrategie. *Zeitschrift für Didaktik der Naturwissenschaften*, 26(1), 73–87. <https://doi.org/10.1007/s40573-020-00113-y>
- Nehring, A., Stiller, J., Nowak, K. H., Upmeyer zu Belzen, A., & Tiemann, R. (2016).

Naturwissenschaftliche Denk- und Arbeitsweisen im Chemieunterricht – eine modellbasierte Video-studie zu Lerngelegenheiten für den Kompetenzbereich der Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 77–96. <https://doi.org/10.1007/s40573-016-0043-2>

NGA Center for Best Practices (NGA Center), Council of Chief State School Officers (CCSSO). (2010). Common Core State Standards Initiative. <http://www.corestandards.org/>

Nickel, H., & Engel, R. K. (2018). Schau genau. *Unterricht Chemie*, 167, 20–24.

Nitz, S., Prechtel, H., & Nerdel, C. (2014). Survey of classroom use of representations: Development, field test and multilevel analysis. *Learning Environments Research*, 17(3), 401–422.

Nowak, K. H., Nehring, A., Tiemann, R., & Upmeyer zu Belzen, A. (2013). Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *Journal of Biological Education*, 47(3), 182–188. <https://doi.org/10.1080/00219266.2013.822747>

Oguz, A., & Yurumezoglu, K. (2007). The Primacy of Observation in Inquiry-Based Science Teaching. In Online Submission. <https://eric.ed.gov/?id=ED498653>

Otteni, M. (2018). Methoden zum Erkunden, Entdecken, Erfinden und Erarbeiten. In *Biologie-Methodik Handbuch für die Sekundarstufe I und II* (S. 76–96). Cornelsen Pädagogik.

Paek, I. (2004). Investigations of differential item functioning: Comparisons among approaches, and extension to a multidimensional context. 1.

Pande, P., & Chandrasekharan, S. (2017). Representational competence: Towards a distributed and embodied cognition account. *Studies in Science Education*, 53(1), 1–43.

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology*, 90(1), 175.

Pfeifer, P. (2002). *Konkrete Fachdidaktik Chemie* Oldenbourg Verlag GmbH. München.

Popper, K. R. (1994). *Logik der Forschung*, 10. Aufl., Tübingen, 3.

Prechtel, M. (2007). „Chemie-Foto-Story“ – Ein Methodenwerkzeug auf dem Weg. 109–127.

Prechtel, M. (2008). Versuchsprotokolle... Einmal anders! Welche Chancen bieten Bildergeschichten? *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung*, 28, 272–276.

Prechtel, M. (2013). Potenziale der sequenziellen Kunst: Bildergeschichten und Comics im naturwissenschaftlichen Unterricht. In U. Hangartner, F. Keller, & D. Oechslin (Hrsg.), *Wissen durch Bilder*. transcript Verlag. <https://doi.org/10.14361/transcript.9783839419830.271>

Quillin, K., & Thomas, S. (2015). Drawing-to-learn: A framework for using drawings to promote model-based reasoning in biology. *CBE-Life Sciences Education*, 14(1), 1–16.

Rasch, G. (1993). *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press.

Reinecke, J., & Pöge, A. (2010). Strukturgleichungsmodelle. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 775–804). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92038-2_29

Ridley, P., & Rogers, A. (2010). *Science, technology, engineering & maths*. University of Brighton.

Roam, D. (2013). *The back of the napkin: Solving problems and selling ideas with pictures*. Marshall Cavendish International (Asia) PTE LTD.

Rost, J. (2004). *Lehrbuch Testtheorie–Testkonstruktion*. Springer.

Roth, W.-M., & McGinn, M. K. (1998). Inscriptions: Toward a theory of representing as social practice. *Review of educational research*, 68(1), 35–59.

- Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Rhetoric and reality in science performance assessments: An update. *Journal of Research in Science Teaching*, 33(10), 1045–1063. [https://doi.org/10.1002/\(SICI\)1098-2736\(199612\)33:10<1045::AID-TEA1>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1098-2736(199612)33:10<1045::AID-TEA1>3.0.CO;2-S)
- Salim, A. S., Brandenburger, M., Mikelskis-Seifert, S., & Schwichow, M. (2020). Intervention zu Variablenkontrollstrategien in der Grundschule. In S. Habig (Hrsg.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen. GDCP Jahrestagung 2019* (Bd. 40). Uni Duisburg-Essen. <http://nbn-resolving.de/urn:nbn:de:0111-pedocs-204454>
- Santelices, M. V., & Wilson, M. (2012). On the relationship between differential item functioning and item difficulty: An issue of methods? Item response theory approach to differential item functioning. *Educational and Psychological Measurement*, 72(1), 5–36.
- Schecker, H., & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12(1), 45–66.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, 8(4), 350–353.
- Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). LGVT 6-12: Lesegeschwindigkeits-und-verständnistest für die Klassen 6-12. Hogrefe.
- Schofield, B. (1989). Science at age 13: A review of APU Survey findings 1980-84. HM Stationery Office.
- Schori, E., Kersten, B., & Abderhalden, C. (2006). Wie reliabel ist der Fragebogen zur Analyse der Pflegedokumentation (IzEP-Dok\copyright) im Instrument zur Erfassung von Pflegesystemen IzEP®. PrInterNet.
- Schreiber, N., Theyßen, H., & Schecker, H. (2014). Diagnostik experimenteller Kompetenz: Kann man Realexperimente durch Simulationen ersetzen? *Zeitschrift für Didaktik der Naturwissenschaften*, 20(1), 161–173. <https://doi.org/10.1007/s40573-014-0017-1>
- Schumacher, S., & Roth, J. (2014). Darstellungskompetenz-Ein Schlüssel zum forschenden Lernen?!. In J. Roth & J. Ames (Hrsg.), *Beiträge zum Mathematikunterricht 2014* (S. 1123–1126). Münster: WTM-Verlag.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Schwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654. <https://doi.org/10.1002/tea.20311>
- Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review*, 39, 37–63. <https://doi.org/10.1016/j.dr.2015.12.001>
- Schwichow, M., & Nehring, A. (2018). Variablenkontrolle beim Experimentieren in Biologie, Chemie und Physik: Höhere Kompetenzausprägungen bei der Anwendung der Variablenkontrollstrategie durch höheres Fachwissen? Empirische Belege aus zwei Studien. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1), 217–233.
- Scott, W. A. (1955). Reliability of Content Analysis: The Case of Nominal Scale Coding. *The Public Opinion Quarterly*, 19(3), 321–325.
- Sodian, B., & Mayer, D. (2013). Entwicklung des wissenschaftlichen Denkens im Vor-und Grundschulalter. In *Handbuch frühkindliche Bildungsforschung* (S. 617–631). Springer.
- Solano-Flores, G., Shavelson, R. J., Ruiz-Primo, M. A., Schults, S. E., Wiley, E. W., & Brown, J. H. (1997). On the Development and Scoring of Classification and Observation Science Performance Assessments. Annual Meeting of the American Educational Research Association, Chicago. <https://eric.ed.gov/?id=ED411314>

- Sommer, K., Wambach-Laicher, J., & Pfeifer, P. (2018). *Konkrete Fachdidaktik Chemie*. Aulis bei Friedrich, Seelze.
- Stecher, B. M., & Klein, S. P. (1996). Performance assessments in science: Hands-on tasks and scoring guides (Bd. 660). Rand Corporation.
- Stein, M., & Power, B. (1996). Putting art on the scientist's palette. *New entries: Learning by writing and drawing*, 59–68.
- Stevens, P. (1978). On the Nuffield Philosophy of Science. *Journal of Philosophy of Education*, 12(1), 99–111. <https://doi.org/10.1111/j.1467-9752.1978.tb00508.x>
- Stieff, M. (2011). When is a molecule three dimensional? A task-specific role for imagistic reasoning in advanced chemistry. *Science Education*, 95(2), 310–336. <https://doi.org/10.1002/sce.20427>
- Stieff, M., Hegarty, M., & Dixon, B. (2010). Alternative Strategies for Spatial Reasoning with Diagrams. In A. K. Goel, M. Jamnik, & N. H. Narayanan (Hrsg.), *Diagrammatic Representation and Inference* (S. 115–127). Springer. https://doi.org/10.1007/978-3-642-14600-8_13
- Stieff, M., & McCombs, M. (2006). Increasing Representational Fluency with Visualization Tools (S. A. Barab, K. E. Hay, & D. T. Hickey, Hrsg.; Bd. 2, S. 730–736). International Society of the Learning Sciences. <https://repository.isls.org/handle/1/3583>
- Stieff, M., & Wilensky, U. (2003). Connected chemistry—Incorporating interactive simulations into the chemistry classroom. *Journal of Science Education and Technology*, 12(3), 285–302.
- Süselbeck, G. (2003). Immer diese Protokolle. *Grundschule*, 35, 59.
- Thomas, J. A., Pedersen, J. E., & Finson, K. (2001). Validating the Draw-A-Science-Teacher-Test Checklist (DASTT-C): Exploring Mental Models and Teacher Beliefs. *Journal of Science Teacher Education*, 12(4), 295–310. <https://doi.org/10.1023/A:1014216328867>
- Tippett, C. D. (2016). What recent research on diagrams suggests about learning with rather than learning from visual representations in science. *International Journal of Science Education*, 38(5), 725–746. <https://doi.org/10.1080/09500693.2016.1158435>
- Tomcin, R., & Reiners, C. S. (2009). Auf malerischem Weg zur Chemie. *CHEMKON*, 16(1), 6–13. <https://doi.org/10.1002/ckon.200910083>
- Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of educational psychology*, 93(1), 129–140.
- Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, 17(4), 285–325.
- Vorholzer, A., von Aufschnaiter, C., & Kirschner, S. (2016). Entwicklung und Erprobung eines Tests zur Erfassung des Verständnisses experimenteller Denk- und Arbeitsweisen. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 25–41.
- Waldrip, B., Prain, V., & Carolan, J. (2010). Using multi-modal representations to improve learning in junior secondary science. *Research in Science Education*, 40(1), 65–80.
- Weinert, F. (2001a). Concept of competence: A conceptual clarification. In *Defining and selecting key competencies* (S. 45–65). Hogrefe & Huber Publishers.
- Weinert, F. (2001b). Vergleichende Leistungsmessung in Schulen—Eine umstrittene Selbstverständlichkeit. In F. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–32). Beltz.
- Wellnitz, N. (2012). *Kompetenzstruktur und-niveaus von Methoden naturwissenschaftlicher Erkenntnisgewinnung*. Logos Verlag Berlin.
- Wellnitz, N., & Mayer, J. (2008). *Evaluation von Kompetenzstruktur und-niveaus zum Beobachten*,

- Vergleichen, Ordnen und Experimentieren. *Erkenntnisweg Biologiedidaktik*, 7, 129–144.
- Wellnitz, N., & Mayer, J. (2013). Erkenntnismethoden in der Biologie–Entwicklung und Evaluation eines Kompetenzmodells. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 315–345.
- Wellnitz, N., & Mayer, J. (2016). Methoden der Erkenntnisgewinnung im Biologieunterricht. Erkenntnisse biologiedidaktischer Forschung. Schwerpunkte und Forschungsstände. *Biologie lernen und lehren*, 61–82.
- Wendell, K. B., & Rogers, C. (2013). Engineering Design-Based Science, Science Content Performance, and Science Attitudes in Elementary School. *Journal of Engineering Education*, 102(4), 513–540. <https://doi.org/10.1002/jee.20026>
- Widden, M., Mackinnon, A., O’shea, T., Wild, R., Shapson, S., Day, E., Pye, I., Moon, B., Cusack, S., & Chin, P. (1992). British Columbia Assessment of Science 1991 Technical Report IV: Context for Science Component. Victoria, Ministry of Education, Multiculturalism and Human Rights of British Columbia.
- Wirth, J., Leutner, D., & Klieme, E. (2005). Problemlösekompetenz—Ökonomisch und zugleich differenziert erfassbar? In E. Klieme, D. Leutner, & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern: Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 73–82). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-85144-4_6
- Wirtz, M. A., & Caspar, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Hogrefe.
- Wirtz, M. A., & Schulz, A. (2012). Experimentieren im mathematisch-naturwissenschaftlichen Unterricht: Schüler lernen wissenschaftlich denken und arbeiten (W. Rieß & B. Barzel, Hrsg.). Waxmann Verlag.
- Wissenschaftliches Konsortium HarmoS. (2008). HarmoS Naturwissenschaften+: Kompetenzmodell und Vorschläge für Bildungsstandards: Wissenschaftlicher Schlussbericht. Büro CLIP.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). ACER ConQuest version 2.0: Generalised item response modelling software.
- Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, 21(2), 179–217. [https://doi.org/10.1016/S0364-0213\(99\)80022-6](https://doi.org/10.1016/S0364-0213(99)80022-6)
- Zöfel, P. (2003). Statistik für Psychologen: Im Klartext (Bd. 7063). Pearson Studium München.

10 Danksagung

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die mich in den letzten Jahren unterstützt und damit zum Gelingen dieser Arbeit beigetragen haben.

Zunächst danke ich Prof. Dr. Maik Walpuski als Erstbetreuer meiner Dissertation für seine Außenperspektive, den fachkundigen und immer vielperspektivischen Rat und die Möglichkeit zur wertvollen Zusammenarbeit mit der Uni Duisburg-Essen.

Mein besonderer Dank gilt Prof. Dr. Christoph Gut, der mich in der gesamten Zeit als mein wissenschaftlicher Mentor unterstützt hat. Bedanken möchte ich mich vor allem für die kritische Auseinandersetzung, die konstruktiven Hilfestellungen und das mir entgegengebrachte Vertrauen.

Herrn Prof. Dr. Micheal Giese danke ich für die Übernahme des Prüfungsvorsitzes.

Der Pädagogischen Hochschule Zürich (PHZH) danke ich für die Förderung durch die Arbeitszeitentlastung (Dissertationsunterstützung) wie auch für die Fördermittel für Material und die studentischen Kodierhelferinnen und -helfer. Dank gebührt auch der Uni Duisburg-Essen für die Nutzung des Scan-Services.

Ich danke Prof. Dr. Roger Keller für die Methodenberatung wie auch für die wertvollen Inputs und Prof. Dr. Maik Philipp für die Schreibberatung und die Auseinandersetzung mit dem Text.

Weiter danke ich allen Schülerinnen und Schülern und deren Lehrpersonen für die Teilnahme an dieser Studie sowie Madleina Niederstein, Lena Rügsegger und Dario Hartmann, die bei der Erhebung und der Kodierung der Daten tatkräftig mitgeholfen haben.

Ebenfalls möchte ich mich bei Dr. Wolfgang Bühner, Prof. Dr. Markus Emden, Jürg Fleischmann und Pitt Hild bedanken, die mich kritisch und mental auch in schwierigen Phasen dieser Arbeit immer unterstützt haben.

Ein weiterer Dank gilt all meinen Kollegen und Kolleginnen an der PHZH, an der Uni Duisburg-Essen und in meinem privaten Umfeld, die mich in den vielen Jahren begleitet haben. Danke für viele schöne Momente im Büro und wertvolle Diskussionen in Kaffeepausen, bei einem Bier und letztlich auch online während der Homeoffice-Phase: Corin Bieri, Markus Brandenburg, Markus Emden, Farida Haidary, Judith Lanka, Thomas Löffel, Urs Schellenberg, Hanspeter Pfirter, Livia Reyes, Zoe Speck,

Josiane Tardent, Anna Trüssel, Vanessa Fischer, Dennis Kirstein, Ines Komor, Adrian Candinas, Cédric Clivaz und Mathias Schnippering.

Ich danke meiner Schwester Ariane Cavelti von Herzen, dass sie immer für mich da war. Ganz besonderer Dank gilt mein Mann Philipp Vetter, der mich ermutigt und unterstützt hat. Vor allem aber danke ich meinen beiden Töchtern Jill und Julie. Ihr zwei bereichert mein Leben.

Danke!

11 Anhang

11.1 Fragenbogen ökologische Validierung

Folgende fachwissenschaftliche Inhalte kommen bei mir im Unterricht vor:

	nie	selten	ab und zu	regelmäßig
1) Chromatographie	☐	☐	☐	☐
2) Farbstift-Chromatographie	☐	☐	☐	☐
3) Phasenübergänge	☐	☐	☐	☐
4) Schmelzen von Eis	☐	☐	☐	☐
5) Einfache chemische Reaktionen	☐	☐	☐	☐
6) Lösen von Brausetabletten	☐	☐	☐	☐
7) CO ₂	☐	☐	☐	☐
8) A) Entstehung von CO ₂ mit Backpulver und Essig	☐	☐	☐	☐
B) Entstehung von CO ₂ mit Backpulver und Zitronensäure	☐	☐	☐	☐
9) Säuren und Basen	☐	☐	☐	☐
10) Indikatoren	☐	☐	☐	☐

Folgende fachwissenschaftliche Inhalte finden sich in Lehrmitteln, die ich für den Unterricht nutze:

		keine	wenige	ab und zu	häufig
11)	Chromatographie.	☐	☐	☐	☐
12)	Farbstift-Chromatographie	☐	☐	☐	☐
13)	Phasenübergänge	☐	☐	☐	☐
14)	Schmelzen von Eis	☐	☐	☐	☐
15)	Einfache chemische Reaktionen	☐	☐	☐	☐
16)	Lösen von Brausetabletten	☐	☐	☐	☐
17)	CO ₂	☐	☐	☐	☐
18)	A) Entstehung von CO ₂ mit Backpulver und Essig	☐	☐	☐	☐
	B) Entstehung von CO ₂ mit Backpulver und Zitronensäure	☐	☐	☐	☐
19)	Säuren und Basen	☐	☐	☐	☐
20)	Indikatoren	☐	☐	☐	☐

Folgende Experimentier-Typen kommen bei mir im Unterricht vor:

		nie	selten	ab und zu	regelmäßig
21)	Beobachtung von Phänomenen	☐	☐	☐	☐
22)	Beobachtung Vergleich zweie Phänomene	☐	☐	☐	☐
23)	Planung eines Experiments zur Erarbeitung von Zusammenhängen	☐	☐	☐	☐

Folgende Experimentier-Typen finden sich in Lehrmitteln, die ich für den Unterricht nutze:

		keine	wenige	ab und zu	häufig
24)	Beobachtung von Phänomenen	☐	☐	☐	☐
25)	Beobachtung Vergleich zweie Phänomene	☐	☐	☐	☐
26)	Planung eines Experiments zur Erarbeitung von Zusammenhängen	☐	☐	☐	☐

Hier folgen ein paar Fragen zu dir

- 27) Ich unterrichte seit folgender Anzahl Jahren
- 28) Ich unterrichte folgende Stufen
- 29) Ich unterrichte folgende Niveaus
- 30) Ich nutze folgende Bücher

11.2 Fragebogen inhaltliche Validierung bezüglich der Qualitätsmerkmale des wissenschaftlichen Skizzierens

Folgende Kriterien sind für wissenschaftliche Skizzen relevant:

		gar nicht	eher nicht	eher ja	unverzicht- bar
1)	deutlich gezeichnete Skizze	☐	☐	☐	☐
2)	wahrnehmbare relevante Qualitätsmerkmale dargestellt (Farbveränderungen, Formveränderungen, Aggregatzustandsänderungen etc.)	☐	☐	☐	☐
3)	Detailtreue	☐	☐	☐	☐
4)	Vollständigkeit	☐	☐	☐	☐
5)	Irrelevantes weggelassen	☐	☐	☐	☐
6)	Alle Vorgänge / Abfolge gezeichnet	☐	☐	☐	☐
7)	Fokus/Relevanz (Unwichtiges weggelassen, Wichtiges im Zentrum)	☐	☐	☐	☐
8)	Abstraktion	☐	☐	☐	☐
9)	angemessene Grösse	☐	☐	☐	☐
10)	relevante Proportionen	☐	☐	☐	☐
11)	Perspektive wechseln	☐	☐	☐	☐
12)	Zoom auf Details	☐	☐	☐	☐

13)	3D- erkenntlich	☐	☐	☐	☐
14)	2D gezeichnet	☐	☐	☐	☐
15)	Symbole (z.B. Pfeile) um Vorgänge darzustellen	☐	☐	☐	☐
16)	Verknüpfungen zwischen zwei Skizzen herstellt	☐	☐	☐	☐
17)	Fachlichkeit (inhaltlich korrekt)	☐	☐	☐	☐
18)	Beschriftung/Legende verwendet	☐	☐	☐	☐
19)	Ergänzung durch Text	☐	☐	☐	☐
20)	Unterscheiden Beobachtung und Erklärung	☐	☐	☐	☐
21)	Vergleich dargestellt	☐	☐	☐	☐
22)	Unterschied erkannt	☐	☐	☐	☐
23)	zwei Phänomene oder Vorgänge verglichen werden, diese nebeneinander gezeichnet	☐	☐	☐	☐
24)	Phänomen wird durch Planung erzeugt	☐	☐	☐	☐
25)	Variablenkontrolle dargestellt	☐	☐	☐	☐

Folgende Kriterien erachte ich auch als relevant

	gar nicht	eher nicht	eher ja	unverzicht- bar
26)	☐	☐	☐	☐
27)	☐	☐	☐	☐

28)

O O O O

Hier folgen zwei Fragen zu Ihnen

29) Ich unterrichte(te) folgender Anzahl Jahren

30) Ich bin seit folgender Anzahl Jahre Fachdidaktiker/Fachdidaktikerin

11.3 Theorieblatt

Regeln des wissenschaftlichen Skizzierens

Wissenschaftliches Skizzieren ist eine Form des Dokumentierens eines Experiments und dessen Planung.

- Schreibe keine Erklärungen auf
- Notiere eine Legende
- Es kommt nicht darauf an, ob du schön zeichnest, sondern was du zeichnest

Durchführung und Beobachtung eines Experiments oder Vergleichs
--

Jemand anderes sollte aus deinen Skizzen das Experiment und das Beobachtete verstehen können.

- Nur zeichnen, was du siehst
- Zeichne jene Vorgänge und Stellen, welche zur Forschungsfrage gehören
- Zeichne wichtige Details vergrößert
- Lasse Unwichtiges weg (z. B. Bunsenbrenner)
- Wechsle die Perspektive (z. B. von der Seite, von oben zeichnen)
- Wenn sich Dinge verändern, kannst du mehrere Skizzen machen
- Wenn zwei Phänomene oder Vorgänge verglichen werden, zeichne diese nebeneinander

Planung eines Experiments

Jemand anderes sollte aus deiner Skizze das Experiment nachmachen können.

- Zeichne den Aufbau des Experiments
- Zeichne alle Geräte und Stoffe, die du brauchst
- Zeichne wie sie angeordnet sind
- Zeichne die Teilschritte des Experiments auf
- Achte darauf, dass man beide kontrollierten Variablen erkennt

11.4 Aufgabenset

Stellvertretend wird die Aufgabenstellung des Items BR_V aufgeführt:

11.4.1 Aufgabenstellung Item BR_V (Vergleich, Brausetablette)

		Name (hier Etikette aufkleben)
--	--	-----------------------------------

Lösen von Smartie und Brausetablette vergleichen

Information

Lina und David sprechen französisch und leben in Lausanne.

Sie verstehen kaum Deutsch.

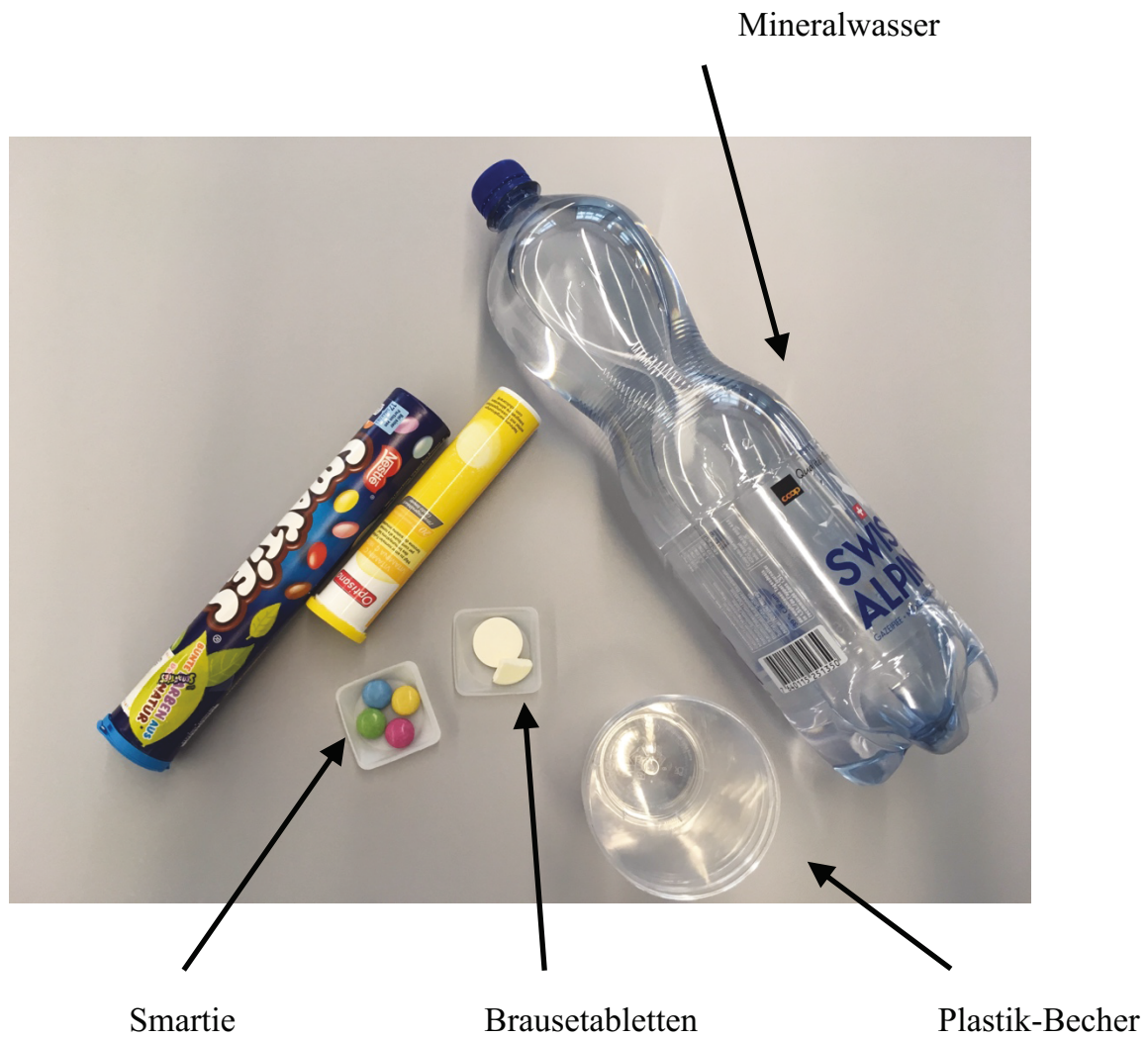
Die beiden haben in einem Experiment herausgefunden, was mit einem Smartie und einer Brausetablette geschieht, wenn diese mit Mineralwasser in Berührung kommen.

Sie haben keine Skizzen angefertigt.

Skizziere für Lina und David.

Beachte dabei: Deine Skizzen sollen von Lina und David verstanden werden, auch wenn sie kein Deutsch verstehen.

Material



Aufgabe

Lina und David haben ihr Experiment gefilmt.

Schau dir den Film an.

Wenn das Wasser das Smartie und die Brausetablette umgibt, verändern sich diese. Beobachte was mit dem Smartie und der Brausetablette geschieht.

Skizziere alle Beobachtungen und das Vorgehen während dem Experimentieren auf. Achte auch auf die Regeln des wissenschaftlichen Skizzierens.

Beantworte mit deinen Skizzen die Frage: Was geschieht mit einem Smartie und einer Brausetablette, wenn diese mit Mineralwasser in Berührung kommen?

Hier ist Platz für deine Skizzen

Legende:

Fragen

Beurteile die folgenden Aussagen auf die Richtigkeit.

Pro Block ist mindestens eine Antwort korrekt, es können aber auch mehrere Antworten richtig sein. Setze pro Block mindestens ein Kreuz mittig in die Kästchen.

Beispiel

- ☒ korrekte Antwort
- ☒ korrekte Antwort
- ☐ falsche Antwort

Wenn du bei einer Frage die Antwort ändern möchtest, fülle das Kästchen mit der nun nicht mehr gültigen Antwort schwarz aus. Das Kreuz setzt du nun in das neue, richtige Kästchen. Solltest du doch die korrigierte Antwort wählen wollen, so kreise die korrigierte Antwort ein.

	Antwort- alternative 1	Antwort- alternative 2	Antwort- alternative 3
Erste, spontane Antwort	☒	☐	☐
Erstes Mal umentschieden (1. Korrektur)	☐	☒	☐
Zweites Mal umentschieden (2. Korrektur)	☒	☐	☐

1. Wenn das Wasser die Brausetablette umgibt, dann...

- ☐ geschieht nichts.
- ☐ sprudelt es.
- ☐ wird das Wasser gasförmig.

2. Wenn das Wasser das Smartie umgibt, dann...

- ☐ geschieht nichts.
- ☐ wird das Wasser gasförmig.
- ☐ sprudelt es.

3. Gas-Blasen beim Smartie...

- ☐ entstehen überall im Wasser.
- ☐ entstehen vor allem beim Smartie.
- ☐ entstehen keine.

4. Gas-Blasen bei der Brausetablette...

- ☐ entstehen keine.
- ☐ entstehen überall im Wasser.
- ☐ entstehen vor allem bei der Brausetablette.

Wenn du fertig bist, kannst du das Sudoku lösen

Fülle die leeren Felder so aus, dass in jeder Zeile und in jeder Spalte und in jedem 3-x-3-Kästchen alle Zahlen von 1 bis 9 nur einmal vorkommen.

		1	3		9	7		
7								6
5	6						1	8
	4		6	9	3		7	
				5				
	7		4	8	2		5	
6	3						8	1
2								9
		4	9		1	3		

4	8	1	3	6	9	7	2	5
7	9	2	5	1	8	4	3	6
5	6	3	2	7	4	9	1	8
1	4	5	6	9	3	8	7	2
3	2	8	1	5	7	6	9	4
9	7	6	4	8	2	1	5	3
6	3	9	7	4	5	2	8	1
2	1	7	8	3	6	5	4	9
8	5	4	9	2	1	3	6	7

Lösung

11.5 Kodiermanual

Stellvertretend wird die Aufgabenstellung des Items BR_V aufgeführt.

11.5.1 Kodiermanual Item BR_V (Vergleich, Brausetablette)

KODIERMANUAL: Diss_BR_V

999 keine Angabe: keine Skizze, kein Text

888 keine eindeutige Angabe: nicht erkennbar

Für das Verständnis der wissenschaftlichen Skizze und die Kodierung der wissenschaftlichen Skizze wird kein Text berücksichtigt, ausser

- der Legende
- Wörter mit Pfeil zu Gegenstand hin (analog der Legende)
- Wörter die der Beschriftung von Becher, Flaschen, Gegenständen dienen (analog der Legende)
- Zahlenwerten wie Temperaturangaben, Mengenangaben, Zahlen zur Beschriftung der Bechergläser

Erklärungen, Begründungen und Beschreibungen von Vorgängen, sowie Fliesstext ist zum Verständnis der wissenschaftlichen Skizze und zur Kodierung der wissenschaftlichen Skizze nicht zu berücksichtigen.

Niveau	Kategorien für «Vergleich»	
	Wissenschaftlichkeit der Skizzen Eindeutigkeit der Skizze (eindeutig interpretierbar): relevante Merkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze eindeutig interpretierbar Idealisierung: Relevanz/Fokus, keine für die Forschungsfrage irrelevanten Merkmale dargestellt Vollständigkeit: Vorgänge/Abläufe dargestellt (Bildergeschichte/Nummern/Verwendung von Pfeilen, um Vorgänge zu zeigen) Abstraktion(sgrad): Abstraktion der Darstellung, keine Fotografie fachliche Korrektheit: relevante Proportionen korrekt, relevante Merkmale der Reaktion korrekt	Unterschied erfasst und Vergleich dargestellt
3	relevante Merkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze grösstenteils ohne zusätzlich Erklärung eindeutig (eindeutig interpretierbar) Idealisierung: Relevanz/Fokus, keine irrelevanten Dinge dargestellt, nur Relevantes, fokussiert, falls Text vorhanden gibt, dieser zusätzliche (nicht zeichenbare) Informationen Vorgänge/Abläufe vollständig dargestellt (Bildergeschichte/Nummern, Verwendung von Pfeilen, um Vorgänge zu zeigen) hoher Abstraktionsgrad grösstenteils fachliche Korrektheit, relevante Proportionen grösstenteils korrekt dargestellt, relevante Merkmale der Reaktion grösstenteils korrekt dargestellt	Vergleich vollständig und korrekt dargestellt (Vergleichsbedingung dargestellt) Relevante Unterschiedsmerkmale vollständig erfasst
2	mindestens relevante Merkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze ohne zusätzlich Erklärung (eindeutig Interpretierbar) nichts bis höchstens wenig Irrelevantes, gar nicht bis höchstens wenig unfokussiert, falls Text vorhanden gibt, dieser zusätzliche (nicht zeichenbare) Informationen Vorgänge/Abläufe mindestens grösstenteils dargestellt (Bildergeschichte/Nummern/Verwendung von Pfeilen, um Vorgänge zu zeigen) mindestens überwiegend abstrahiert mindestens teilweise fachlich korrekt, relevante Merkmale der Reaktion mindestens teilweise korrekt dargestellt (nur kleine Fehler oder etwas Fehlendes)	Vergleich vollständig korrekt dargestellt Relevante Unterschiedsmerkmale mindestens teilweise erfasst
1	mindestens grösstenteils relevante Merkmale zum Erstellen des Phänomens dargestellt, grösstenteils Phänomen dargestellt, Skizze mindestens teilweise ohne zusätzlich Erklärung eindeutig (grösstenteils eindeutig interpretierbar) mindestens teilweise Irrelevantes, teilweise unfokussiert, falls Text vorhanden sagen Text und Skizze dasselbe (redundant) Vorgang des Reaktionsverlaufs mindestens teilweise gezeichnet (Pfeile ansatzweise verwendet) teilweise Bildabfolgen mindestens teilweise abstrahiert (eher Fotografie-ähnlich) mindestens grösstenteils fachlich nicht korrekt (fehlerhaft, fehlendes)	Vergleichsvariablen erkannt Vergleich unvollständig und mindestens nicht vollkommen korrekt dargestellt Relevante Unterschiedsmerkmale nicht oder mindestens ansatzweise erfasst

0	nicht oder ansatzweise relevante Merkmale zum Erstellen des Phänomens dargestellt, Phänomen dargestellt, Skizze nicht oder ansatzweise eindeutig ohne zusätzliche Erklärung (ansatzweise eindeutig interpretierbar) nicht oder mindestens teilweise Irrelevantes, unfokussiert, falls Text vorhanden kann dieser in eine Skizze überführt werden Reaktionsverlauf nicht oder nur teilweise dargestellt nicht oder mindestens ansatzweise abstrahiert (Fotografie-ähnlich) nicht oder nur mindestens ansatzweise korrekt (viele fachliche Fehler und Fehlendes)	Vergleichsvariablen nicht oder mindestens teilweise erkannt, aber nicht vollständig dargestellt Relevante Unterschiedsmerkmale nicht erfasst
---	---	--

Indikatoren Wissenschaftliches Skizzieren

A1) Sind die relevanten Merkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Merkmale dargestellt? (Eindeutigkeit der Skizze)

W11) Ist der Vorgang des Experiments erkennbar und nachvollziehbar nur mit der Legende? (Wasser, Brausetablette, Smarties vorhanden, Brausetablette und Smarties im Wasser)

W12) Sind Gasblasen dargestellt (als Kreise oder Punkte)?

A2) Es werden nur relevante Merkmale fokussiert (Idealisierung)

W21) Sind die wesentlichen Materialien dargestellt: Becher, Wasser, Smarties und Brausetablette?

W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) die Brausetablette-Dose, Smarties-Dose und Wasserflasche ii) das Schälchen mit Brausetablette (weggelassen!) iii) wurde alles andere weggelassen (Hände etc.)

W23) Ist neben der Legende ((Mineral)Wasser, Brausetablette, Smarties, Becher) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?

A3) Sind alle relevanten Merkmale der Reaktion dargestellt? (Vollständigkeit)

W31) Gibt es eine Abfolge des ganzen Vorgangs (mehrerer Bilder, Pfeile, Nummer als Abfolge)?

W32) Wurde die Reaktion (Auflösen der Brausetablette/Smarties) in mehreren Bildern dargestellt?

W33) Ist erkenntlich: i) nach Reaktion keiner Brausetablette mehr ii) eingefärbte Wasser nach Reaktion, iii) nach der Reaktion keinen Gasblasen mehr iv) Smarties nach der Reaktion noch vorhanden?

A4) Sind die relevanten Merkmale genügend abstrahiert? (Abstraktion)

W41) Wurde abstrahiert gezeichnet i) Wasserglas (2D/ pseudo 3D?) ii) Brausetablette (Kreis/Oval)?

W42) Wurde das Auflösen der Brausetablette/Smarties dargestellt durch Kreise/Streiche, Punkte?

A5) Sind die relevanten Merkmale fachliche korrekt dargestellt? (Korrektheit)

W51) Gehen die Blasen von Brausetablette/Smarties aus (nicht überall Blasen)?

W52) Ist die Ausbreitung der Blasen korrekt dargestellt: i) Smarties direkt nach oben, Brausetablette deutlich breiter als beim Smartie)

W53) Ist die Grösse der Blasen korrekt dargestellt i) Blasen beim Smartie kleiner als bei der Brausetablette?

Indikatoren Vergleich dargestellt, Unterschied erfasst

V1) Sind die zu vergleichenden Sachverhalte nebeneinander dargestellt? (Brausetablette und Smarties)

V2) Ist erkenntlich, dass beide unter denselben Bedingungen behandelt werden (gleich viel Wasser im Becher, Brausetablette und das Smartie werden ins (Mineral)wasser gegeben)?

V3) Ist erkenntlich, dass sowohl beim Smartie wie auch bei der Brausetablette es zu einer Reaktion kommt (Blasenentstehung bei beiden)?

V4) Die relevanten Unterschiede des Vergleichs erfasst worden?

i) unterschiedliche Ausbreitung der Gasblasen

iii) unterschiedliche Blasengröße

V5) Sind die relevanten Unterschiede bezüglich der Blasenorts deutlich dargestellt?

i) Gasblasen beim Smartie sehr begrenzt oberhalb des Smarties

ii) Gasblasen bei der Brausetablette deutlich breiter sich ausbreiten dargestellt

V6) Sind die relevanten Unterschiede bezüglich deutlich der Blasengröße dargestellt?

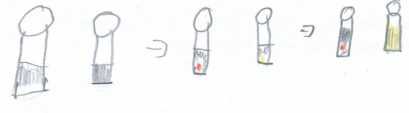
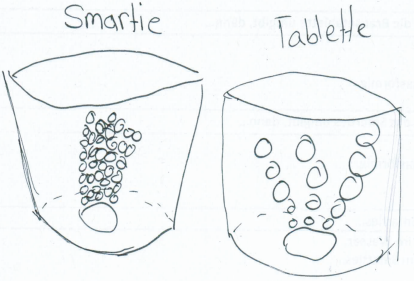
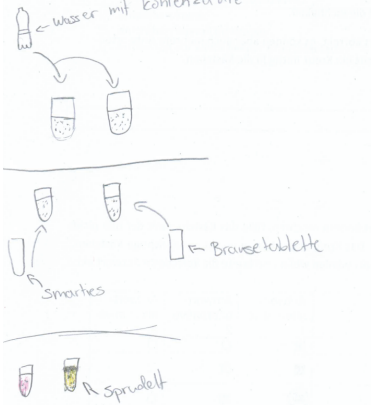
i) Blasen beim Smartie deutlich kleiner

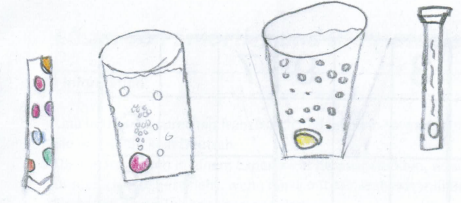
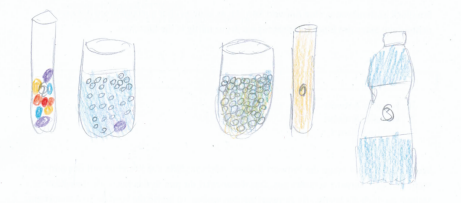
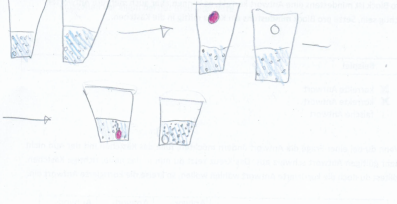
V7) Sind die relevanten Unterschiede bezüglich des Gegenstands deutlich dargestellt?

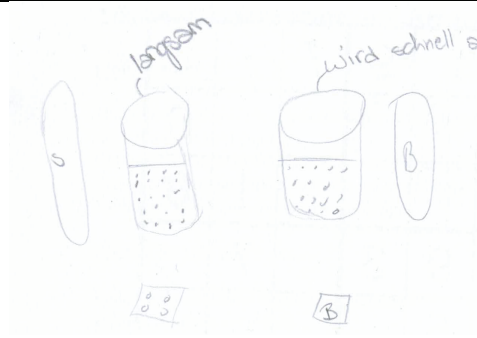
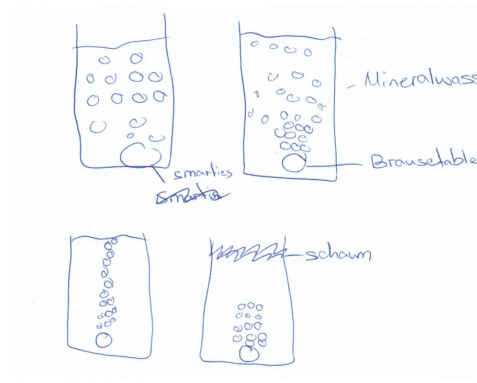
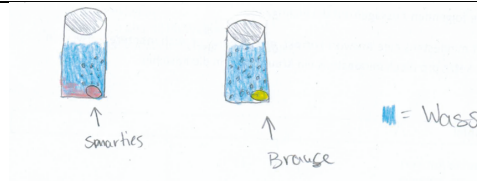
i) Smarties löst sich nicht auf/Brausetablette schon

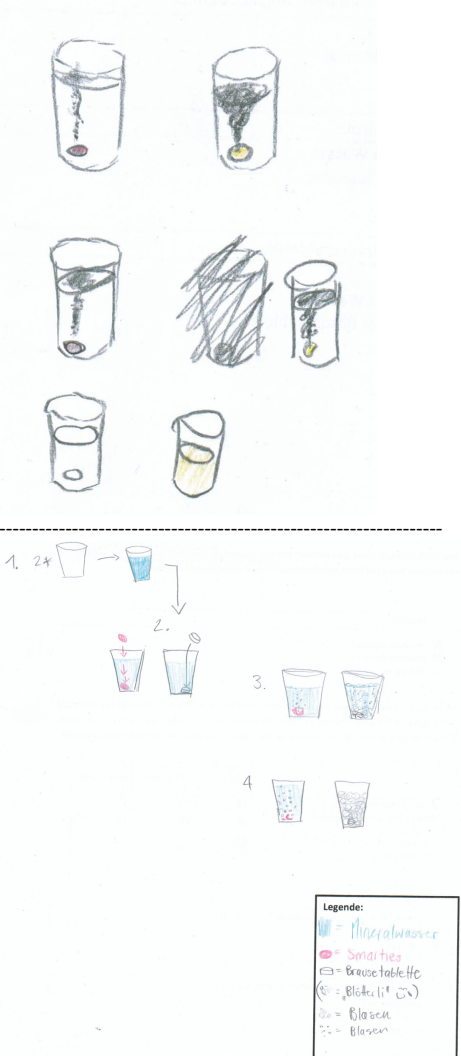
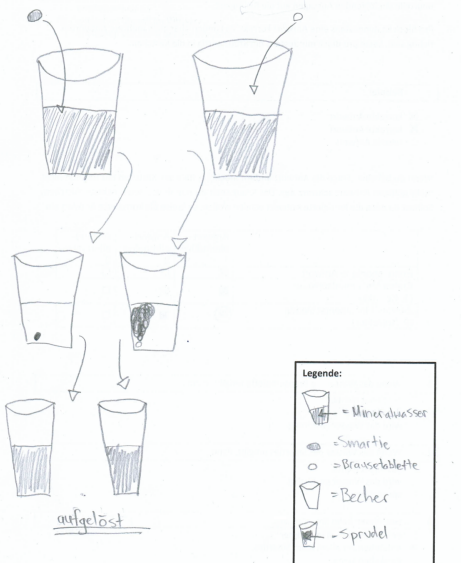
Schwellenwerte (ohne W53, ohne V6)

	Wissenschaftlichkeit der Skizze	Vergleich
2 → 3	$[\geq 2, \geq 3, \geq 3, \geq 2, \geq 1]$	$[\geq 5]$
1 → 2	$[\geq 1, \geq 2, \geq 2, \geq 2, \geq 0]$	$[\geq 4]$
0 → 1	$[\geq 0, \geq 2, \geq 1, \geq 1, \geq 0]$	$[\geq 2]$

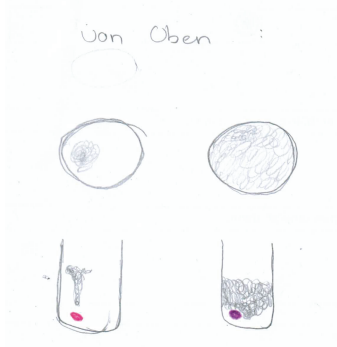
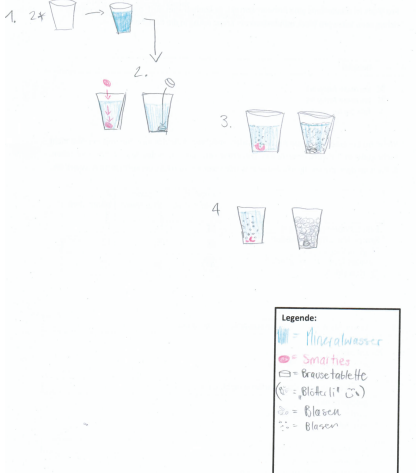
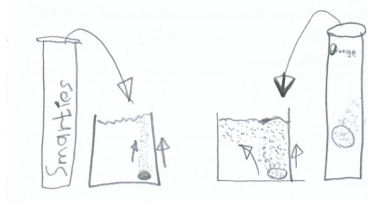
Indikator	Beschreibung	Ankerbeispiele
W11) Ist der Vorgang des Experiments erkennbar und nachvollziehbar nur mit der Legende? (Wasser, Brausetablette, Smarties vorhanden, Brausetablette und Smarties im Wasser)	Skizze gibt den Vorgang des Experiments wieder, man erkennt, dass die Brausetablette und Smarties ins Wasser gegeben wurden oder im Wasser. Oder wird ins Wasser gegeben (mit Pfeil dargestellt)	 Fehler: nicht erkennbar was vor sich geht – Grösse viel zu klein (Originalgrösse)
W12) Sind Gasblasen dargestellt (als Kreise oder Punkte oder Striche)?	Die Blasen des Mineralwassers sind falsch (diese sind gleichmässig verteilt) => 0 falls eine Brausetablette/Smarties im Glas ist, sind es die Gasblasen => 1 Falls es klar die Blasen des Mineralwassers sind (sehr gleichmässig verteilt, obwohl die Tablette im Wasser ist) => 0	 Fehler: Blasen des Mineralwassers 
W21) Sind die wesentlichen Materialien dargestellt: Becher, Wasser, Smarties und Brausetablette?	Die Materialien müssen gezeichnet sein (nur in der Legende reicht nicht)	

W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) die Brausetablette-Dose, Smarties-Dose und Wasserflasche ii) das Schälchen mit Brausetablette (weggelassen!) iii) wurde alles andere weggelassen (Hände etc.)	Ist die Hälfte oder mehr korrekt => 1	 Fehler: Schälchen wurden dargestellt.  Fehler: Smartiedose zu genau, Brausetablette-Dose ok  Fehler: Mineralwasserflasche zu genau abgezeichnet.
W23) Ist neben der Legende ((Mineral)Wasser, Brausetablette, Smarties, Becher) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?	Beispiele: Gefärbtes Wasser, nachher nur Wiederholung des Titels (Vergleich von..) =>1	 Fehler: redundanter Text in der Legende Legende: 7. Mineralwasser / Sprudelwasser in zwei Bechern geben. 7. In einem Becher den Smarties in den anderen eine Brausetablette 5. Vorher was geschieht.

		 Zusätzliche Infos im Text  Das Wort «Schaum»
W31) Gibt es eine Abfolge des ganzen Vorgangs (mehrerer Bilder, Pfeile, Nummer als Abfolge)?	mehr als eine Skizze im total (Skizzen müssen aber den Vorgang darstellen) (auch ohne Pfeil oder Nummern ok)	 Fehler: nur jeweils ein Bild 

W32) Wurde die Reaktion (Auflösen der Brausetablette/Smarties) in mehreren Bildern dargestellt?	mehr als eine Skizze der Reaktion (hier Auflösen der Brausetablette/Smarties) Ein Bild mit Brausetablette im Wasser + ein mit Reaktion (bei Brausetablette Verfärbung oder Blasen) => 1	
W33) Ist erkenntlich: i) nach Reaktion keiner Brausetablette mehr ii) eingefärbte Wasser nach Reaktion, iii) nach der Reaktion keinen Gasblasen mehr iv) Smarties nach der Reaktion noch vorhanden?	Ist die Hälfte oder mehr korrekt => 1	

		Fehler: da Smarties sich auch auflöst, mehr Abfolge beim Auflösen
W41) Wurde abstrahiert gezeichnet i) Wasserglas (2D/ pseudo 3D?) ii) Brausetablette (Kreis/Oval)?	Ist die Hälfte oder mehr korrekt => 1 Pseudo 3D = nur oben einen 3D-Ring, wenn auch unten der Rand 3D oder der Wasserspiegel 3D oder beides => 0	
W42) Wurde das Auflösen der Brausetablette/Smarties dargestellt durch Kreise/Streiche, Punkte?	Brausetablette muss vorhanden sein: wenn W12 => 0 dann W42 => 0	

W51) Gehen die Blasen von Brausetablette/Smarties aus (nicht überall Blasen)?	wenn W12 => 0 dann W51 => 0 wenn nur Smarties oder Brausetablette korrekt dann => 1	
W52) Ist die Ausbreitung der Blasen korrekt dargestellt: i) Smarties direkt nach oben, ii) Brausetablette deutlich breiter als beim Smartie	wenn W12 => 0 dann W52 => 0 wenn Blasen nicht oberhalb der Brausetablette gezeichnet => 0 Deutlicher Unterschied in der Breite erkennbar, sonst => 0 Erkennbar was Smarties und was Brausetablette ist sonst => 0	 

W53) Ist die Grösse der Blasen korrekt dargestellt i) Blasen beim Smartie kleiner als bei der Brausetablette?

wenn W12 => 0 dann W53 => 0

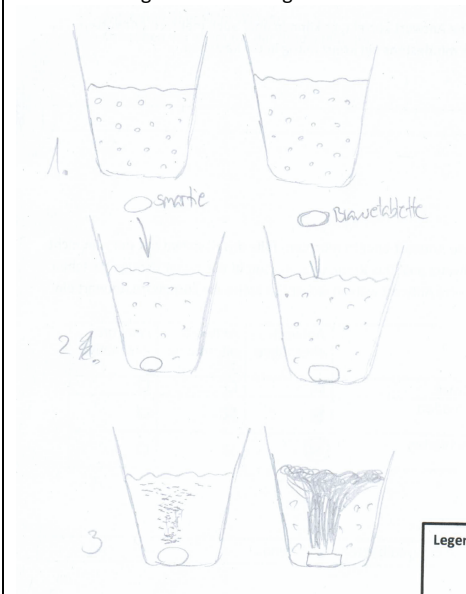
Wenn Darstellung der Blasen deutlich anderes ist => 1

Sowohl beim Smartie wie bei der Brausetablette müssen Blasen gezeichnet sein sonst => 0

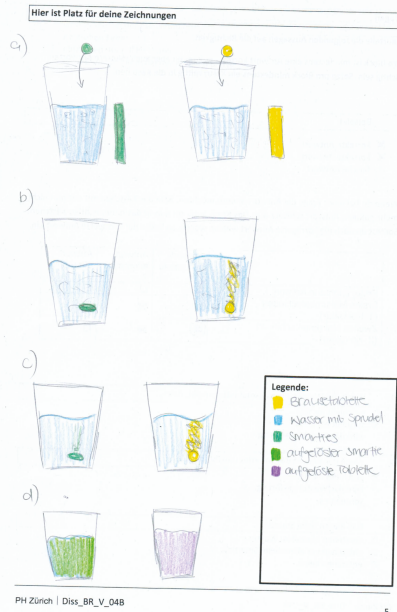
Erkennbar was Smarties und was Brausetablette ist sonst => 0



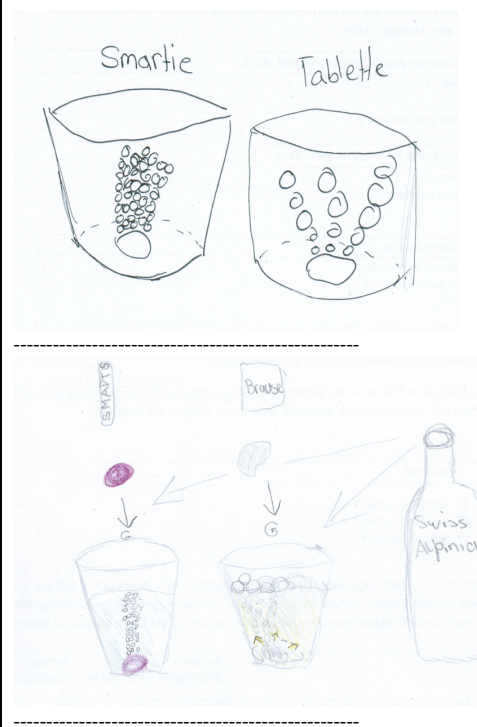
Fehler: Blaesgrösse nicht dargestellt



Darstellung der Blasen deutlich anders

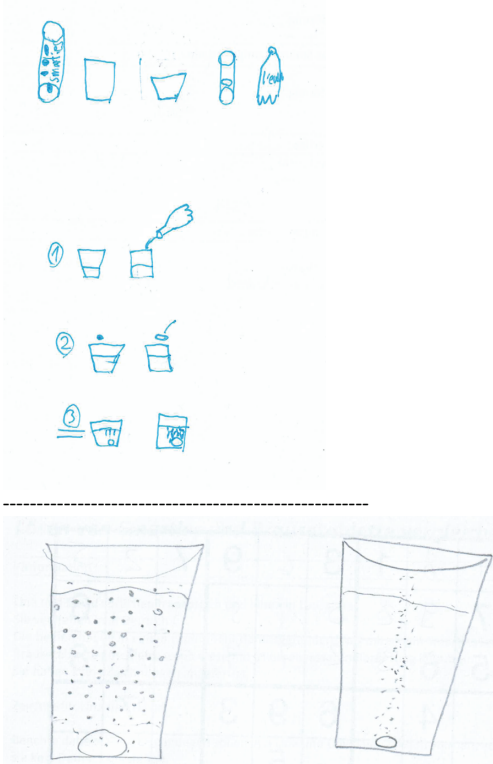
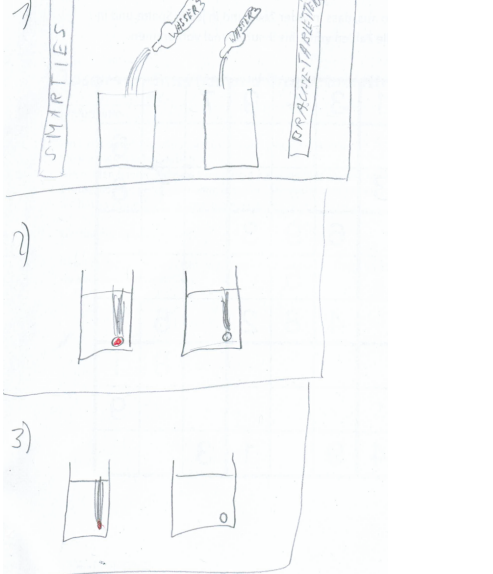


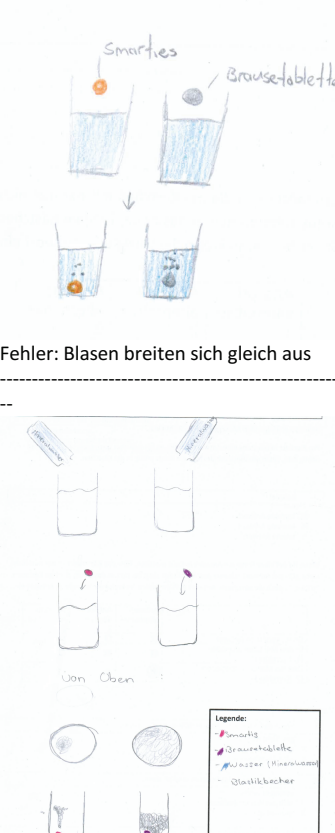
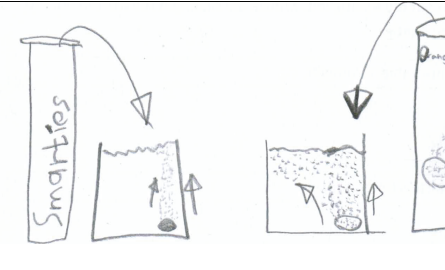
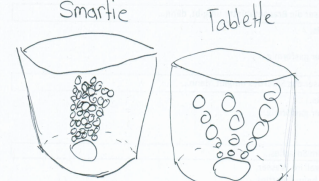
Darstellung der Blasen deutlich anders



		 Darstellung deutlich anders.  Darstellung deutlich anders – Blasengröße klar ersichtlich Legende: Becher mit Mineralwasser Brausetablette Smarties
--	--	---

Indikator	Beschreibung	Ankerbeispiele
V1) Sind die zu vergleichenden Sachverhalte nebeneinander dargestellt? (Brausetablette und Smarties)	beide Becher nebeneinander/untereinander dargestellt. Es muss etwas in den Bechern enthalten sein.	

V2) Ist erkenntlich, dass beide unter denselben Bedingungen behandelt werden (gleich viel Wasser im Becher, Brausetablette und des Smarties werden ins (Mineral)wasser gegeben)?	<i>Ist die Hälfte oder mehr korrekt => 1</i> <i>Falls V1 => 0, V2=> 0</i> <i>Wenn Wasserpegel nicht eingezeichnet (Grenze Blasen oder Schaum gilt auch) => 0</i> <i>Falls V1 => 0, V2=> 0</i> <i>(Brausetablette und Smarties müssen ins Wasser gegeben werden oder sich im Wasser befinden sonst => 0)</i>	
V3) Ist erkenntlich, dass sowohl beim Smartie wie auch bei der Brausetablette es zu einer Reaktion kommt (Blasenentstehung bei beiden)?	<i>Reaktion = Blasen</i> <i>Beide Reaktionen müssen dargestellt sein sonst => 0</i> <i>Falls V1 => 0, V3=> 0</i>	

V4) Die die relevanten Unterschiede des Vergleichs erfasst worden? i) unterschiedliche Ausbreitung der Gasblasen ii) unterschiedliche Blasengrösse	Ist die Hälfte oder mehr korrekt => 1 Wenn nur eines gezeichnet (entweder Smarties oder Brausetablette) oder nur i) oder ii) für Smarties und Brausetablette korrekt => 1 Es muss erkennbar sein was Brausetablette und was Smarties ist. Zu i) Blasen müssen unterschiedlich gross sein. Auch grössere Blasen beim Smarties => 1	 Fehler: Blasen breiten sich gleich aus
V5) Sind die relevanten Unterschiede bezüglich der Blasenorts deutlich dargestellt? i) Gasblasen beim Smartie sehr begrenzt oberhalb dem Smartie ii) Gasblasen bei der Brausetablette deutlich breiter sich ausbreitend dargestellt	Wenn nicht Gasblasen bei Smarties und Brausetablette gezeichnet => 0 Es muss erkennbar sein was Brausetablette und was Smarties ist.	
V6) Sind die relevanten Unterschiede bezüglich deutlich der Blasengrösse dargestellt? i) Blasen beim Smartie deutlich kleiner	Wenn nicht Gasblasen bei Smarties und Brausetablette gezeichnet => 0 Wenn Darstellung der Blasen deutlich anderes ist => 1 Es muss erkennbar sein was Brausetablette und was Smarties ist.	

V7) Sind die relevanten Unterschiede bezüglich des Gegenstands deutlich dargestellt? i) Smartie löst sich nicht auf/Brausetablette schon	<i>beide müssen erfüllt sein => 1</i> <i>Es muss erkennbar sein was Brausetablette und was Smarties ist.</i>	
---	---	--

11.6 Indikatoren

11.6.1 Indikatoren BR_B

A1) Sind die relevanten Objektmerkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt? (Eindeutigkeit der Zeichnung)

W11) Ist der Vorgang des Experiments erkennbar und nachvollziehbar nur mit der Legende? (Wasser, Brausetablette, Brausetablette im Wasser)

W12) Sind Gasblasen dargestellt (als Kreise, Punkte, Striche)?

A2) Es werden nur relevante Objektmerkmale fokussiert (Idealisierung)

W21) Sind die wesentlichen Materialien dargestellt: Becher, Wasser und Brausetablette?

W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) die Brausetablette-Dose und Wasserflasche ii) das Schälchen mit Brausetablette (weggelassen!) iii) wurde alles andere weggelassen (Hände etc.)

W23) Ist neben der Legende (Wasser, Brausetablette, Becher) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?

A3) Sind alle relevanten Objektmerkmale der Reaktion dargestellt? (Vollständigkeit)

W31) Gibt es eine Abfolge des ganzen Vorgangs (mehrerer Bilder, Pfeile, Nummer als Abfolge)?

W32) Wurde die Reaktion (Lösen der Brausetablette) in mehreren Bildern dargestellt?

W33) Ist erkenntlich: i) nach Reaktion keiner Brausetablette mehr ii) eingefärbte Wasser nach Reaktion, iii) nach der Reaktion keinen Gasblasen mehr?

A4) Sind die relevanten Objektmerkmale genügend abstrahiert? (Abstraktion)

W41) Wurde abstrahiert gezeichnet i) Wasserglas (2D/ pseudo 3D?) ii) Brausetablette (Kreis/Oval/ Rechteck)?

W42) Wurde das Lösen der Brausetablette dargestellt durch Kreise/Striche, Punkte?

A5) Sind die relevanten Objektmerkmale fachliche korrekt dargestellt? (Korrektheit)

W51) Gehen die Blasen von Brausetablette aus (nicht überall Blasen)?

W52) Sind Bewegung der Brausetablette dargestellt (hoch und runter mit Pfeilen oder in mehreren Bildern)?

11.6.2 Indikatoren EI_B

A1) Sind die relevanten Objektmerkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt? (Eindeutigkeit der Zeichnung)

W11) Ist der Vorgang des Experiments erkennbar und nachvollziehbar nur mit der Legende? (Eiswürfel, Salz, Salz auf Eiswürfel)

W12) Ist der Schmelzvorgang dargestellt (als Linie/Einbuchtung oder/und kleiner werden des Eiswürfels)?

A2) Es werden nur relevante Objektmerkmale fokussiert (Idealisierung)

W21) Sind die wesentlichen Materialien dargestellt: Eiswürfel, Salz?

W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) Salzgefäß, Uhrglas ii) das Schälchen mit Eiswürfel (weggelassen!) iii) wurde alles andere weggelassen (Hände etc.)

W23) Ist neben der Legende (Eiswürfel, Salz, Uhrglas) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?

A3) Sind alle relevanten Objektmerkmale der Reaktion dargestellt? (Vollständigkeit)

W31) Gibt es eine Abfolge des ganzen Vorgangs (mehrerer Bilder, Pfeile, Nummer als Abfolge)?

W32) Wurde die Reaktion (Schmelzen des Eises) in mehreren Bildern dargestellt?

W33) Ist erkenntlich: i) Verkleinerung des Eiswürfels ii) die Entstehung von Wasser durch das Schmelzen?

A4) Sind die relevanten Objektmerkmale genügend abstrahiert? (Abstraktion)

W41) Wurde abstrahiert gezeichnet i) Salzgefäß (2D/ pseudo 3D oder weggelassen?)

ii) Eiswürfel (Viereck/Oval)? (falls Salzgefäß weggelassen und sehr detailliert gezeichnet =0)

- A5) Sind die relevanten Objektmerkmale fachliche korrekt dargestellt? (Korrektheit)
 W51) Sind Furchen/Einbuchtungen des Schmelzvorgangs mit dem Salz erkennbar?
 W52) Ist dargestellt, dass das Eis mit Salz schneller schmilzt?

11.6.3 Indikatoren SB_B

- A1) Sind die relevanten Objektmerkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt? (Eindeutigkeit der Zeichnung)
 W11) Ist der Vorgang des Experiments erkennbar und nachvollziehbar nur mit der Legende? (Universal)indikator, (verdünnte) Essigsäure, (verdünnte) Natronlauge, Zutropfen/Zugeben der Stoffe nacheinander)
 W12) Ist ein Farbveränderungen dargestellt?
 A2) Es werden nur relevante Objektmerkmale fokussiert (Idealisierung)
 W21) Sind die wesentlichen Materialien dargestellt: RG, (Universal)indikator, (verdünnte) Essigsäure, (verdünnte) Natronlauge?
 W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) Flaschen von (Universal)indikator, (verdünnte) Essigsäure, (verdünnte) Natronlauge, ii) Pipette iii) wurde alles andere weggelassen (RG-Ständer etc.)
 W23) Ist neben der Legende (RG, (Universal)indikator, (verdünnte) Essigsäure, (verdünnte) Natronlauge, Pipette) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?
 A3) Sind alle relevanten Objektmerkmale der Reaktion dargestellt? (Vollständigkeit)
 W31) Gibt es eine Abfolge des ganzen Vorgangs (mehrerer Bilder, Pfeile, Nummer als Abfolge)?
 W32) Wurde die Reaktion (Farbveränderung) in mehreren Bildern dargestellt?
 W33) Ist erkenntlich: i) die Lösung mit Universalindikator rot wurde ii) die Lösung mit Natronlauge die Farbe wechselt
 A4) Sind die relevanten Objektmerkmale genügend abstrahiert? (Abstraktion)
 W41) Wurde abstrahiert gezeichnet i) RG (2D/ pseudo 3D?) ii) Pipette?
 W42) Wurde das Zutropfte dargestellt durch Tropfen/Punkte/Striche oder Pfeil?
 A5) Sind die relevanten Objektmerkmale fachliche korrekt dargestellt? (Korrektheit)
 W51) Ist der Farbwechsel von Rot zu Blauviolett klar dargestellt?
 W52) Ist die Schlierenbildung klar dargestellt?

11.6.4 Indikatoren BR_V

A1) Sind die relevanten Objektmerkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt? (Eindeutigkeit der Zeichnung)

W11) Ist der Vorgang des Experiments erkennbar und nachvollziehbar nur mit der Legende? (Wasser, Brausetablette, Smartie vorhanden, Brausetablette und Smartie im Wasser)

W12) Sind Gasblasen dargestellt (als Kreise oder Punkte)?

A2) Es werden nur relevante Objektmerkmale fokussiert (Idealisierung)

W21) Sind die wesentlichen Materialien dargestellt: Becher, Wasser, Smartie und Brausetablette?

W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) die Brausetablette-Dose, Smartie-Dose und Wasserflasche ii) das Schälchen mit Brausetablette (weggelassen!) iii) wurde alles andere weggelassen (Hände etc.)

W23) Ist neben der Legende ((Mineral)Wasser, Brausetablette, Smartie, Becher) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?

A3) Sind alle relevanten Objektmerkmale der Reaktion dargestellt? (Vollständigkeit)

W31) Gibt es eine Abfolge des ganzen Vorgangs (mehrerer Bilder, Pfeile, Nummer als Abfolge)?

W32) Wurde die Reaktion (Lösen der Brausetablette/Smartie) in mehreren Bildern dargestellt?

W33) Ist erkenntlich: i) nach Reaktion keiner Brausetablette mehr ii) eingefärbte Wasser nach Reaktion, iii) nach der Reaktion keinen Gasblasen mehr iv) Smartie nach der Reaktion noch vorhanden?

A4) Sind die relevanten Objektmerkmale genügend abstrahiert? (Abstraktion)

W41) Wurde abstrahiert gezeichnet i) Wasserglas (2D/ pseudo 3D?) ii) Brausetablette (Kreis/Oval)?

W42) Wurde das Lösen der Brausetablette/Smartie dargestellt durch Kreise/Streiche, Punkte?

A5) Sind die relevanten Objektmerkmale fachliche korrekt dargestellt? (Korrektheit)

W51) Gehen die Blasen von Brausetablette/Smartie aus (nicht überall Blasen)?

W52) Ist die Ausbreitung der Blasen korrekt dargestellt: i) Smartie direkt nach oben, Brausetablette deutlich breiter als beim Smartie)

W53) Ist die Grösse der Blasen korrekt dargestellt i) Blasen beim Smartie kleiner als bei der Brausetablette?

V1) Sind die zu vergleichenden Sachverhalte nebeneinander dargestellt? (Brausetablette und Smartie)

V2) Ist erkenntlich, dass beide unter denselben Bedingungen behandelt werden (gleich viel Wasser im Becher, Brausetablette und das Smartie werden ins (Mineral)wasser gegeben)?

V3) Ist erkenntlich, dass sowohl beim Smartie wie auch bei der Brausetablette es zu einer Reaktion kommt (Blasenentstehung bei beiden)?

V4) Die die relevanten Unterschiede des Vergleichs erfasst worden?

i) unterschiedliche Ausbreitung der Gasblasen

iii) unterschiedliche Blasengröße

V5) Sind die relevanten Unterschiede bezüglich der Blasenorts deutlich dargestellt?

i) Gasblasen beim Smartie sehr begrenzt oberhalb dem Smartie

ii) Gasblasen bei der Brausetablette deutlich breiter sich ausbreiten dargestellt

V6) Sind die relevanten Unterschiede bezüglich deutlich der Blasengröße dargestellt?

i) Blasen beim Smartie deutlich kleiner

V7) Sind die relevanten Unterschiede bezüglich des Gegenstands deutlich dargestellt?

i) Smartie löst sich nicht auf/Brausetablette schon

11.6.5 Indikatoren EI_V

A1) Sind die relevanten Objektmerkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt? (Eindeutigkeit der Zeichnung)

W11) Ist der Vorgang des Experiments erkennbar und nachvollziehbar nur mit der Legende? (Eiswürfel, Salz- und Süsswasser, Eiswürfel im Wasser)

W12) Ist der Schmelzvorgang dargestellt (Eiswürfel wird kleiner)?

A2) Es werden nur relevante Objektmerkmale fokussiert (Idealisierung)

W21) Sind die wesentlichen Materialien dargestellt: Eiswürfel, Salz- Süsswasser, Becher?

W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) Salzwasserflasche, Süsswasserflasche ii) das Schälchen mit Eiswürfel (weggelassen!) iii) wurde alles andere weggelassen (Hände etc.)

W23) Ist neben der Legende (Eiswürfel, Salz- Süßwasser) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?

A3) Sind alle relevanten Objektmerkmale der Reaktion dargestellt? (Vollständigkeit)

W31) Gibt es eine Abfolge des ganzen Vorgangs (mehrerer Bilder, Pfeile, Nummer als Abfolge)?

W32) Wurde die Reaktion (Schmelzen des Eiswürfels) in mehreren Bildern dargestellt?

W33) Ist erkenntlich beide Eiswürfel werden kleiner.

W34) Ist erkenntlich eingefärbtes Wasser?

A4) Sind die relevanten Objektmerkmale genügend abstrahiert? (Abstraktion)

W41) Wurde abstrahiert gezeichnet i) Becher (2D/ pseudo 3D?) ii) Eiswürfel (Viereck/Oval)?

W42) Wurden die Einfärbungen des Wassers dargestellt durch Schattierung/Farbe, Punkte?

A5) Sind die relevanten Objektmerkmale fachliche korrekt dargestellt? (Korrektheit)

W51) Ist die Einfärbung des Wassers korrekt dargestellt i) farbige Ring beim Salzwasser, ii) eingefärbte Wasser beim Süßwasser überall?

W52) Ist der Schmelzvorgang korrekt dargestellt? (beim Süßwasser deutlich schneller)

W53) Wurde der Eiswürfel sowohl beim Salz- wie auch beim Süßwasser gezeichnet?

W53b) Ragt der Eiswürfel beim Salzwasser mehr aus dem Wasser?

V1) Sind die zu vergleichenden Sachverhalte nebeneinander dargestellt? (Salz- und Süßwasser?)

V2) Ist erkenntlich, dass beide unter denselben Bedingungen behandelt werden (gleich viel Wasser im Becher, Eiswürfel werden ins (Salz- und Süß)wasser gegeben)?

V3) Ist erkenntlich, dass sowohl beim Salz- wie auch beim Süßwasser es zu einer Reaktion kommt (Schmelzen des Eiswürfels bei beiden, Einfärbung des Wassers)

V4) Die die relevanten Unterschiede des Vergleichs erfasst worden?

i) unterschiedliche rascher Schmelzvorgang (Grösse des Eiswürfels)

ii) unterschiedliche Einfärbung des Wassers

iii) unterschiedlicher Auftrieb des Eiswürfels (wie weit er aus dem Wasser ragt)

V5) Sind die relevanten Unterschiede bezüglich des Eiswürfels deutlich dargestellt?

- i) rascheres Schmelzen beim Süsswasser ((Grösse des Eiswürfels)
- ii) Eiswürfel ragt bei Salzwasser mehr aus dem Wasser heraus (unterschiedlicher Auftrieb des Eiswürfels)
- V6) Sind die relevanten Unterschiede bezüglich des Einfärbens deutlich dargestellt?
 - i) farbiger Ring beim Salzwasser
- V7) Sind die relevanten Unterschiede bezüglich des Einfärbens deutlich dargestellt?
 - ii) Einfärbung vollständig beim Süsswasser

11.6.6 Indikatoren SB_V

- A1) Sind die relevanten Objektmerkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt? (Eindeutigkeit der Zeichnung)
 - W11) Ist der Vorgang des Experiments erkennbar und nachvollziehbar nur mit der Legende? (Universal)indikator, (verdünnte) Essigsäure, (verdünnte) Natronlauge, Sodalösung Zutropfen/Zugeben der Stoffe nacheinander)
 - W12) Ist ein Farbveränderungen dargestellt?
 - W13) Sind Gasblasen dargestellt?
- A2) Es werden nur relevante Objektmerkmale fokussiert (Idealisierung)
 - W21) Sind die wesentlichen Materialien dargestellt: RG, (Universal)indikator, (verdünnte) Essigsäure, (verdünnte) Natronlauge, Sodalösung?
 - W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) Flaschen von (Universal)indikator, (verdünnte) Essigsäure, (verdünnte) Natronlauge, Sodalösung, ii) Pipette iii) wurde alles andere weggelassen (RG-Ständer etc.)
 - W23) Ist neben der Legende (RG, (Universal)indikator, (verdünnte) Essigsäure, (verdünnte) Natronlauge, Sodalösung, Pipette) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?
- A3) Sind alle relevanten Objektmerkmale der Reaktion dargestellt? (Vollständigkeit)
 - W31) Gibt es eine Abfolge des ganzen Vorgangs (mehrerer Bilder, Pfeile, Nummer als Abfolge)?
 - W32) Wurde die Reaktion (Farbveränderung) in mehreren Bildern dargestellt?
 - W33) Ist erkenntlich: i) die Lösung mit Universalindikator rot wurde ii) die Lösung mit Natronlauge die Farbe wechselt iii) die Lösung mit der Sodalösung die Farbe wechselt, iv) bei der Lösung mit der Sodalösung Gasblasen entstehen.
- A4) Sind die relevanten Objektmerkmale genügend abstrahiert? (Abstraktion)

- W41) Wurde abstrahiert gezeichnet i) RG (2D/ pseudo 3D?) ii) Pipette?
- W42) Wurde das Zugetropfte dargestellt durch Tropfen/Punkte/Striche oder Pfeile?
- W43) Wurde der Regenbogen abstrahiert durch Farbverläufe?
- A5) Sind die relevanten Objektmerkmale fachliche korrekt dargestellt? (Korrektheit)
- W51) Ist der Farbwechsel von Rot zu Regenbogen klar dargestellt?
- W52) Sind die Gasblasen korrekt dargestellt i) bei der Sodalösung, ii) aus der Zwischenschicht ausgehend?
- V1) Sind die zu vergleichenden Sachverhalte nebeneinander dargestellt? (Sodalösung und Natronlauge?)
- V2) Ist erkenntlich, dass beide unter denselben Bedingungen behandelt werden (gleiche Stoffe, die zugegeben werden (Universalindikator, Essigsäure), Sodalösung oder Natronlauge werden zu getropft, RG gleich hoch gefüllt)
- V3) Ist erkenntlich, dass sowohl bei Sodalösung wie auch der Natronlauge es zu einer Reaktion kommt (Entstehung eines Regenbogens.)
- V4) Die relevanten Unterschiede des Vergleichs erfasst worden?
- i) Gasblasenentstehung nur bei einem RG
- V5) Sind die relevanten Unterschiede deutlich dargestellt?
- i) Gasblasen bei der Sodalösung
- ii) Gasblasen aus der Zwischenschicht heraus?

11.6.7 Indikatoren BR_E

- A1) Sind die relevanten Objektmerkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt? (Eindeutigkeit der Zeichnung)
- W11) Ist der Vorgang der Planung erkennbar und nachvollziehbar nur mit der Legende? (Wasser unterschiedlicher Temperatur, Brausetablette, Brausetablette im Wasser)
- A2) Es werden nur relevante Objektmerkmale fokussiert (Idealisierung)
- W21) Sind die wesentlichen Materialien dargestellt: Becher, warmes und kaltes Wasser und Brausetablette?
- W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) die Brausetablette-Dose, Thermosflasche und Wasserflasche ii) das Schälchen mit Brausetablette (weggelassen!) iii) wurde alles andere weggelassen (Hände etc.)

W23) Ist neben der Legende (Wasser, Thermosflasche, Brausetablette, Becher) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?

A3) Sind alle relevanten Objektmerkmale der Planung dargestellt? (Vollständigkeit)

W31) Ist erkenntlich, dass Variable 1 kontrolliert wurde (Wassertemperatur variiert)?

W32) Ist erkenntlich, dass Variable 2 kontrolliert wurde (Wassermenge variiert)?

W33) Wurde die Reaktion der Brausetablette nicht gezeichnet?

A4) Sind die relevanten Objektmerkmale genügend abstrahiert? (Abstraktion)

W41) Wurde abstrahiert gezeichnet i) Wasserglas (2D/ pseudo 3D?) ii) Brausetablette (Kreis/Oval)? iii) Messzylinder ohne Strichlein

W42) Wurde die Variable 1 (Wassertemperatur) abstrahiert (blau/rot, Wasserdampf, 20°/40°)?

W43) Wurde die Variable 2 (Wassermenge) abstrahiert? (keine Angaben in ml)?

A5) Sind die relevanten Objektmerkmale fachliche korrekt dargestellt? (Korrektheit)

W51) Ist die Kontrolle der Variablen 1 (Wassertemperatur) dargestellt? (verschiedene Wassertemperaturen)

W52) Ist die Kontrolle der Variablen 2 (Wassermenge) dargestellt? (verschiedene Wassermengen)

W53) Wurden 4 Becher dargestellt?

V1) Ist erkenntlich, dass mit unterschiedlicher Wassermenge gearbeitet wird?

V2) Ist erkenntlich, dass mit unterschiedlicher Wassertemperatur gearbeitet wird?

V3) Ist erkenntlich, dass es sich um Becher mit unterschiedlichen Wassermengen handelt?

V4) Ist erkenntlich, dass es sich um Becher mit unterschiedlichen Wassertemperaturen handelt?

V5) Ist erkenntlich, dass jeweils 2x2 Becher gleich viel Wasser darin haben?

V6) Ist erkenntlich, dass jeweils 2x2 Becher (unterschiedlicher Wassermenge) die gleiche Temperatur haben? (20°/40°, kalt/warm, rot/blau)

11.6.8 Indikatoren EI_E

A1) Sind die relevanten Objektmerkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt? (Eindeutigkeit der Zeichnung)

W11) Ist der Vorgang der Planung erkennbar und nachvollziehbar nur mit der Legende? (Salz- und Süßwasser, Eiswürfel, Eiswürfel im Wasser)

A2) Es werden nur relevante Objektmerkmale fokussiert (Idealisierung)

W21) Sind die wesentlichen Materialien dargestellt: Becher, Salz- und Süßwasser und Eiswürfel?

W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) Salz- und Süßwasserflasche ii) das Schälchen mit Eiswürfel (weggelassen!) iii) wurde alles andere weggelassen (Hände etc.)

W23) Ist neben der Legende (Becher, Salz- und Süßwasser und Eiswürfel) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?

A3) Sind alle relevanten Objektmerkmale der Planung dargestellt? (Vollständigkeit)

W31) Ist erkenntlich, dass Variable 1 kontrolliert wurde (Salzgehalt (Salz-Süßwasser) variiert)?

W32) Ist erkenntlich, dass Variable 2 kontrolliert wurde (Wassermenge variiert)?

W33) Wurde die Reaktion mit dem Eiswürfel nicht gezeichnet?

A4) Sind die relevanten Objektmerkmale genügend abstrahiert? (Abstraktion)

W41) Wurde abstrahiert gezeichnet i) Wasserglas (2D/ pseudo 3D?) ii) Eiswürfel (Viereck/Oval)? iii) Messzylinder ohne Strichlein

W42) Wurde die Variable 1 (Salzgehalt) abstrahiert (Salz/Süßwasser, Salzdose gezeichnet oder nicht, Punkte/Strich im Wasser für Salz)?

W43) Wurde die Variable 2 (Wassermenge) abstrahiert? (keine Angaben in ml)?

A5) Sind die relevanten Objektmerkmale fachliche korrekt dargestellt? (Korrektheit)

W51) Ist die Kontrolle der Variablen 1 (Salzgehalt) dargestellt? (Salz- und Süßwasser)

W52) Ist die Kontrolle der Variablen 2 (Wassermenge) dargestellt? (verschiedene Wassermengen)

W53) Wurden 4 Becher dargestellt?

V1) Ist erkenntlich, dass mit unterschiedlichen Salzgehalten gearbeitet wird?

V2) Ist erkenntlich, dass mit unterschiedlichen Wassermengen gearbeitet wird?

V3) Ist erkenntlich, dass es sich um Becher mit unterschiedlichen Salzgehalten handelt?

V4) Ist erkenntlich, dass es sich um Becher mit unterschiedlichen Wassermengen handelt?

V5) Ist erkenntlich, dass jeweils 2x2 Becher (unterschiedlicher Wassermenge) den gleichen Salzgehalt haben?

V6) Ist erkenntlich, dass jeweils 2x2 Becher gleich viel Wasser darin haben?

11.6.9 Indikatoren SB_E

A1) Sind die relevanten Objektmerkmale zum Erstellen des Phänomens dargestellt, sind alle relevanten Objektmerkmale dargestellt? (Eindeutigkeit der Zeichnung)

W11) Ist der Vorgang der Planung erkennbar und nachvollziehbar nur mit der Legende? ((Schwarz)tee, Zugabe von Seife und Zitronensaft zum Schwarztee)

A2) Es werden nur relevante Objektmerkmale fokussiert (Idealisierung)

W21) Sind die wesentlichen Materialien dargestellt: Becher, (Schwarz)tee, Seife und Zitronensaft?

W22) Wurde weggelassen oder falls gezeichnet idealisiert? i) die Seifen-Gefäß, Zitronensaftflasche, Tee flasche, ii) Pipette, iii) wurde alles andere weggelassen (Hände etc.)

W23) Ist neben der Legende (Becher, (Schwarz)tee, Seife, Zitronensaft) kein zusätzlicher Text vorhanden oder falls Text vorhanden ist, gibt dieser zusätzliche Infos und ist nicht redundant?

A3) Sind alle relevanten Objektmerkmale der Planung dargestellt? (Vollständigkeit)

W31) Ist erkenntlich, dass Variable 1 kontrolliert wurde (Zugabe zu Schwarztee) variiert)?

W32) Ist erkenntlich, dass Variable 2 kontrolliert wurde (Tee-Menge variiert)?

W33) Wurde die Reaktion mit Seife und Zitronensaft nicht gezeichnet?

A4) Sind die relevanten Objektmerkmale genügend abstrahiert? (Abstraktion)

W41) Wurde abstrahiert gezeichnet i) Tee-Becher (2D/ pseudo 3D?) ii) Messzylinder ohne Strichlein

W42) Wurde die Variable 1 (Zugabe zu Schwarztee) abstrahiert i) Seife (ev. mit Händewaschen)? iii) Zitronensaft (ev. mit Zitrone)?

W43) Wurde die Variable 2 (Teemenge) abstrahiert? (keine Angaben in ml)?

A5) Sind die relevanten Objektmerkmale fachliche korrekt dargestellt? (Korrektheit)

W51) Ist die Kontrolle der Variablen 1 (Zugabe zu Schwarztee) dargestellt? (Seife und Zitronensaft)

W52) Ist die Kontrolle der Variablen 2 (Teemenge) dargestellt? (verschiedene Teemengen)

W53) Wurden 4 Becher dargestellt

V1) Ist erkenntlich, dass mit unterschiedlicher Teemenge gearbeitet wird?

V2) Ist erkenntlich, dass mit unterschiedlichen Stoffen als Zugabe gearbeitet wird?

V3) Ist erkenntlich, dass es sich um Becher mit unterschiedlichen Teemengen handelt?

V4) Ist erkenntlich, dass es sich um Becher mit unterschiedlichen Stoffen als Zugabe handelt?

V5) Ist erkenntlich, dass jeweils 2x2 Becher gleich viel Tee darin haben?

V6) Ist erkenntlich, dass jeweils in 2x2 Becher (unterschiedlicher Teemenge) das gleiche zugegeben wurde? (Seife/Zitronensaft)

11.7 Interrater Auswertung (Übereinstimmung/Cohens Kappa/Gwet AC1)

Aufgabentyp	Kontext	Indikator	Übereinstimmung	Cohens κ	Gwet AC1	Inferenz	Schiefe	Mittelwert	Standardabweichung
Beobachtung	Brau-setablette	W11	1	1	1	0	-4.659	0.959	0.199
		W12	1	1	1	0	-2.566	0.893	0.309
		W21	1	1	1	0	-6.177	0.975	0.155
		W22	0.950	0.474	0.945	0	-4.945	0.963	0.189
		W23	0.905	0.8813	0.914	0	-1.214	0.758	0.429
		W31	0.975	0.918	0.964	0	-1.569	0.807	0.395
		W32	0.975	0.9484	0.9515	0	-0.066	0.516	0.501
		W33	0.925	0.8462	0.8545	0	0.718	0.332	0.472
		W41	0.975	0.90826	0.9656	0	-2.794	0.906	0.293
		W42	1	1	1	0	-2.498	0.889	0.314
		W51	0.85	0.6522	0.7414	1	-0.216	0.553	0.498
		W52	1	1	1	0	2.370	0.119	0.324

Aufgabentyp	Kontext	Indikator	Übereinstimmung	Cohens κ	Gwet AC1	Inferenz	Schiefe	Mittelwert	Standardabweichung
Beobachtung	Eis schmelzen	W11	1	1	1	0	-4.313	0.953	0.212
		W12	0.85	0.68338	0.7175	1	-0.468	0.613	0.488
		W21	1	nicht def	1	0	-5.420	0.969	0.174
		W22	1	1	1	0	-5.420	0.969	0.174
		W23	0.925	0.8077	0.8776	0	-1.314	0.773	0.419
		W31	0.975	0.8438	0.9702	0	-2.394	0.883	0.322

		W32	0.975	0.918	0.964	0	-1.236	0.762	0.427
		W33	0.9575	0.9355	0.9592	1	-0.854	0.695	0.461
		W41	0.975	0.9083	0.9657	0	-2.972	0.914	0.281
		W51	0.875	0.541	0.8282	1	1.750	0.172	0.378
		W52	0.875	0.6575	0.804	1	1.485	0.203	0.403

Aufgabentyp	Kontext	Indikator	Übereinstimmung	Cohens κ	Gwet AC1	Inferenz	Schiefe	Mittelwert	Standardabweichung
Beobachtung	Säuren und Basen	W11	0.9286	0.8531	0.861	0	0.511	0.377	0.486
		W12	0.95238	0.8931	0.9143	0	-0.492	0.619	0.487
		W21	0.9762	0.9091	0.9678	0	-0.819	0.689	0.464
		W22	0.9254	0.86875	0.9253	0	-0.547	0.631	0.483
		W23	0.9523	0.9523	0.9425	0	-2.096	0.861	0.347
		W31	0.9762	0.6557	0.9744	0	-1.440	0.791	0.407
		W32	0.9285	0.8363	0.8733	0	-0.475	0.615	0.488
		W33	0.9285	0.9285	0.8634	1	-0.334	0.582	0.494
		W41	1	nicht def	1	0	-5.682	0.971	0.167
		W42	0.9762	0.9417	0.9605	0	-0.529	0.627	0.485
		W51	0.9	0.7059	0.8486	1	0.948	0.287	0.453
		W52	0.881	0.7368	0.7827	1	0.621	0.453	0.479

Aufgabentyp	Kontext	Indikator	Übereinstimmung	Cohens κ	Gwet AC1	Inferenz	Schiefe	Mittelwert	Standardabweichung
Vergleich	Brau-setabelle	W11	0.9762	0	0.9756	0	-3.830	0.943	0.233
		W12	0.9762	0.8955	0.9692	1	-1.710	0.824	0.395
		W21	1	nicht def	1	0	-8.906	0.988	0.110
		W22	0.9524	0.4815	0.9476	0	-2.970	0.914	0.280
		W23	0.9524	0.86875	0.925	0	-1.040	0.730	0.444
		W31	0.9524	0.86875	0.9253	0	-0.283	0.570	0.495
		W32	0.881	0.7619	0.7631	0	0.334	0.418	0.493
		W33	0.9048	0.81	0.8095	0	0.738	0.328	0.469
		W41	0.9762	0	0.9756	0	-3.170	0.922	0.268
		W42	0.9762	0.8955	0.9692	1	-1.673	0.820	0.384
		W51	0.8571	0.6111	0.7758	1	-1.063	0.734	0.499
		W52	0.9524	0.9048	0.9048	0	0.099	0.475	0.500
		W53	0.8333	0.1878	0.7939	1	2.370	0.119	0.324

		V1	1	nicht def	1	0	nicht def	1	0
		V2	0.9524	0	0.9233	0	-2.794	0.906	0.292
		V3	0.9762	0.91892	0.9663	1	-1.471	0.795	0.404
		V4	0.881	0.763	0.762	1	0.033	0.492	0.500
		V5	0.881	0.761	0.7631	1	0.300	0.426	0.495
		V6	0.8333	0.1878	0.7939	1	2.714	0.098	0.298
		V7	0.9048	0.80556	0.81333	0	1.322	0.225	0.418

Aufgabentyp	Kontext	Indikator	Übereinstimmung	Cohens κ	Gwet AC1	Inferenz	Schiefte	Mittelwert	Standardabweichung
Vergleich	Eis schmelzen	W11	1	1	1	0	-3.665	0.938	0.241
		W12	0.975	0.9441	0.9548	1	-0.175	0.543	0.498
		W21	0.975	0.6552	0.9731	0	4.185	0.951	0.217
		W22	0.925	0.375	0.915	0	-2.706	0.901	0.298
		W23	0.925	0.8324	0.8643	0	-1.316	0.774	0.418
		W31	0.925	0.79425	0.8824	0	-0.634	0.650	0.477
		W32	0.95	0.89333	0.9059	1	-0.108	0.527	0.499
		W33	0.95	0.8961	0.90384	1	0.124	0.469	0.499
		W34	0.95	0.9	0.9	0	0.397	0.403	0.491
		W41	0.925	0	0.9192	0	-2.787	0.905	0.293
		W42	0.95	0.9	0.9	0	0.414	0.399	0.490
		W51	0.9	0.8	0.8005	0	0.712	0.333	0.471
		W52	0.875	0.7093	0.7808	1	0.898	0.296	0.457
		W53	1	1	1	0	-2.872	0.909	0.287
		W53b	1	1	1	1	3.665	0.062	0.241
		V1	1	1	1	0	-4.185	0.951	0.217
		V2	0.95	0.7243	0.939	0	-2.706	0.901	0.298
		V3	0.875	0.734	0.7676	1	-0.075	0.519	0.500
		V4	0.95	0.7727	0.936	1	2.040	0.144	0.351
		V5	0.9	0.7342	0.84	1	0.813	0.313	0.464
		V6	0.95	0.84375	0.9264	0	1.562	0.193	0.395
		V7	0.9	0.7917	0.8077	0	1.234	0.239	0.426

Aufgabentyp	Kontext	Indikator	Übereinstimmung	Cohens κ	Gwet AC1	Inferenz	Schiefte	Mittelwert	Standardabweichung
Vergleich		W11	0.925	0.8477	0.8523	0	0.246	0.439	0.496

	Säuren und Basen	W12	0.925	0.80645	0.8776	1	-0.674	0.659	0.474
		W13	1	1	1	0	-1.062	0.733	0.442
		W21	0.9	0.8	0.8	0	0.246	0.439	0.496
		W22	0.975	0.931	0.9608	0	-0.601	0.643	0.479
		W23	0.9	0.7143	0.8464	0	-1.205	0.757	0.429
		W31	0.975	0.9083	0.9657	0	-0.548	0.631	0.482
		W32	0.9	0.8826	0.8165	0	0.295	0.427	0.495
		W33	0.925	0.8065	0.8776	0	-0.583	0.639	0.48
		W41	0.925	0	0.9744	0	-6.324	0.976	0.152
		W42	0.9225	0.8378	0.8605	0	-0.198	0.549	0.498
		W43	0.925	0.8065	0.8776	1	-0.601	0.643	0.479
		W51	0.825	0.6517	0.6502	0	0.769	0.322	0.467
		W52	1	1	1	0	0.071	0.482	0.5
		V1	1	nicht def	1	0	-6.324	0.976	0.152
		V2	0.925	0.362	0.915	0	-2.330	0.878	0.327
		V3	0.925	0.8266	0.8685	1	-0.513	0.624	0.485
		V4	1	1	1	0	-0.868	0.698	0.459
		V5	0.95	0.89796	0.9022	0	0.548	0.369	0.482

Aufgabentyp	Kontext	Indikator	Übereinstimmung	Cohens κ	Gwet AC1	Inferenz	Schiefte	Mittelwert	Standardabweichung
Experimentieren	Brau- setab- lette	W11	1	1	1	0	-1.308	0.773	0.419
		W21	1	1	1	0	-2.221	0.871	0.336
		W22	0.95	0.7231	0.939	0	-3.373	0.929	0.256
		W23	0.95	0.7222	0.939	0	-2.448	0.886	0.317
		W31	1	1	1	0	3.155	0.922	0.269
		W32	1	1	1	0	-1.983	0.851	0.356
		W33	1	1	1	0	-0.055	0.514	0.500
		W41	0.975	0	0.974	0	-9.110	0.988	0.108
		W42	1	1	1	0	-3.155	0.922	0.269
		W43	0.975	0.9419	0.9662	0	-0.619	0.647	0.478
		W51	0.975	0.6562	0.9731	0	-2.073	0.859	0.348
		W52	1	1	1	0	-1.509	0.800	0.400
		W53	0.9575	0	0.9745	0	-6.324	0.976	0.152
		V1	0.95	0.4737	0.9448	0	-2.027	0.855	0.352
		V2	1	1	1	0	-3.057	0.918	0.275

		V3	0.975	0.7872	0.9717	0	-1.898	0.843	0.364
		V4	0.95	0.4805	0.9446	0	-2.448	0.886	0.317
		V5	1	1	1	0	-1.672	0.820	0.385
		V6	0.925	0.375	0.915	0	-1.205	0.757	0.429

Aufgabentyp	Kontext	Indikator	Übereinstimmung	Cohens κ	Gwet AC1	Inferenz	Schiefe	Mittelwert	Standardabweichung
Experimentieren	Eis schmelzen	W11	0.9048	0.7375	0.8505	0	-1.087	0.738	0.440
		W21	0.9762	0.8757	0.97056	0	-2.047	0.857	0.351
		W22	0.9286	0.7586	0.8989	0	-2.096	0.861	0.346
		W23	0.9524	0.475	0.9476	0	-4.945	0.963	0.188
		W31	1	1	1	0	-2.566	0.893	0.309
		W32	0.9524	-0.024	0.95006	0	-2.254	0.873	0.333
		W33	1	nicht def	1	0	-8.906	0.988	0.110
		W41	1	nicht def	1	0	-5.682	0.971	0.167
		W42	1	1	1	0	-2.566	0.893	0.309
		W43	0.905	0.7565	0.8447	0	-0.758	0.676	0.468
		W51	0.9048	0.7181	0.8564	0	-1.214	0.758	0.428
		W52	0.9762	0.78788	0.9762	0	-1.710	0.824	0.381
		W53	1	nicht def	1	0	-8.906	0.988	0.110
		V1	1	1	1	0	-2.638	0.898	0.303
		V2	0.95238	-0.02	0.9501	0	-2.638	0.877	0.328
		V3	1	1	1	0	-2.311	0.836	0.370
		V4	0.9524	0.475	0.9476	0	-1.955	0.848	0.359
		V5	0.881	0.6847	0.8095	0	-1.016	0.725	0.446
		V6	0.9286	0.3636	0.9196	0	-1.955	0.848	0.359

Aufgabentyp	Kontext	Indikator	Übereinstimmung	Cohens κ	Gwet AC1	Inferenz	Schiefe	Mittelwert	Standardabweichung
Experimentieren	Säuren und Basen	W11	0.95	0.89899	0.901	0	0.033	0.492	0.500
		W21	0.975	0.931	0.9608	0	-1.226	0.760	0.427
		W22	1	nicht def	1	0	-5.707	0.972	0.166
		W23	0.925	0.513	0.9107	0	-3.083	0.919	0.273
		W31	0.925	0.8171	0.8729	0	-0.791	0.683	0.465
		W32	1	1	1	0	-2.325	0.878	0.327

		W33	0.95	0.4805	0.9448	0	-6.840	0.980	0.141
		W41	0.975	0	0.9744	0	-7.697	0.984	0.126
		W42	0.9	0.6875	0.8529	0	-1.307	0.772	0.419
		W43	0.95	0.8746	0.9169	0	-1.006	0.724	0.447
		W51	0.9	0.79798	0.80198	0	-0.115	0.528	0.499
		W52	1	1	1	0	-2.061	0.858	0.349
		W53	1	nicht def	1	0	-4.433	0.955	0.207
		V1	1	1	1	0	-2.325	0.878	0.327
		V2	0.925	0.6875	0.9017	0	-1.364	0.780	0.414
		V3	1	1	1	0	-2.268	0.874	0.332
		V4	0.95	0.886	0.9109	0	-0.673	0.659	0.474
		V5	0.975	0.875	0.9687	0	-2.109	0.862	0.345
		V6	0.925	0.8492	0.8508	0	-0.033	0.508	0.500

Bisher erschienene Bände der Reihe
Studien zum Physik- und Chemielernen

ISSN 1614-8967

Vollständige Übersicht auf unserer Website



<https://www.logos-verlag.de/spcl>

Aktuelle Bände

- 300 Amany Annaggar (2020): A Design Framework for Video Game-Based Gamification Elements to Assess Problem-solving Competence in Chemistry Education
ISBN 978-3-8325-5150-6 52.00 EUR
- 301 Alexander Engl (2020): CHEMIE PUR – Unterrichten in der Natur. Entwicklung und Evaluation eines kontextorientierten Unterrichtskonzepts im Bereich Outdoor Education zur Änderung der Einstellung zu „Chemie und Natur“
ISBN 978-3-8325-5174-2 59.00 EUR (open access)
- 302 Christin Marie Sajons (2020): Kognitive und motivationale Dynamik in Schülerlaboren. Kontextualisierung, Problemorientierung und Autonomieunterstützung der didaktischen Struktur analysieren und weiterentwickeln
ISBN 978-3-8325-5155-1 56.00 EUR (open access)
- 303 Philipp Bitzenbauer (2020): Quantenoptik an Schulen. Studie im Mixed-Methods Design zur Evaluation des Erlanger Unterrichtskonzepts zur Quantenoptik
ISBN 978-3-8325-5123-0 59.00 EUR (open access)
- 304 Malte Ubben (2020): Typisierung des Verständnisses mentaler Modelle mittels empirischer Datenerhebung am Beispiel der Quantenphysik
ISBN 978-3-8325-5181-0 43.50 EUR (open access)
- 305 Wiebke Hinrike Kuske-Janßen (2020): Sprachlicher Umgang mit Formeln von LehrerInnen im Physikunterricht am Beispiel des elektrischen Widerstandes in Klassenstufe 8
ISBN 978-3-8325-5183-4 47.50 EUR (open access)
- 306 Kai Bliesmer (2020): Physik der Küste für außerschulische Lernorte. Eine Didaktische Rekonstruktion
ISBN 978-3-8325-5190-2 58.00 EUR (open access)
- 307 Nikola Schild (2021): Eignung von domänenspezifischen Studieneingangsvariablen als Prädiktoren für Studienerfolg im Fach und Lehramt Physik
ISBN 978-3-8325-5226-8 42.00 EUR (open access)

- 308 Daniel Aeverbeck (2021): Zum Studienerfolg in der Studieneingangsphase des Chemie-
studiums. Der Einfluss kognitiver und affektiv-motivationaler Variablen
ISBN 978-3-8325-5227-5 51.00 EUR
- 309 Martina Strübe (2021): Modelle und Experimente im Chemieunterricht. Eine Video-
studie zum fachspezifischen Lehrerwissen und -handeln
ISBN 978-3-8325-5245-9 45.50 EUR
- 310 Wolfgang Becker (2021): Auswirkungen unterschiedlicher experimenteller Repräsen-
tationen auf den Kenntnisstand bei Grundschulkindern
ISBN 978-3-8325-5255-8 50.00 EUR
- 311 Marvin Rost (2021): Modelle als Mittel der Erkenntnisgewinnung im Chemieunter-
richt der Sekundarstufe I. Entwicklung und quantitative Dimensionalitätsanalyse eines
Testinstruments aus epistemologischer Perspektive
ISBN 978-3-8325-5256-5 44.00 EUR (open access)
- 312 Christina Kobl (2021): Förderung und Erfassung der Reflexionskompetenz im Fach
Chemie
ISBN 978-3-8325-5259-6 41.00 EUR (open access)
- 313 Ann-Kathrin Beretz (2021): Diagnostische Prozesse von Studierenden des Lehramts.
eine Videostudie in den Fächern Physik und Mathematik
ISBN 978-3-8325-5288-6 45.00 EUR (open access)
- 314 Judith Breuer (2021): Implementierung fachdidaktischer Innovationen durch das An-
gebot materialgestützter Unterrichtskonzeptionen. Fallanalysen zum Nutzungsverhal-
ten von Lehrkräften am Beispiel des Münchener Lehrgangs zur Quantenmechanik
ISBN 978-3-8325-5293-0 50.50 EUR (open access)
- 315 Michaela Oettle (2021): Modellierung des Fachwissens von Lehrkräften in der Teil-
chenphysik. Eine Delphi-Studie
ISBN 978-3-8325-5305-0 57.50 EUR (open access)
- 316 Volker Brüggemann (2021): Entwicklung und Pilotierung eines adaptiven Multistage-
Tests zur Kompetenzerfassung im Bereich naturwissenschaftlichen Denkens
ISBN 978-3-8325-5331-9 40.00 EUR (open access)
- 317 Stefan Müller (2021): Die Vorläufigkeit und soziokulturelle Eingebundenheit natur-
wissenschaftlicher Erkenntnisse. Kritische Reflexion, empirische Befunde und fachdi-
daktische Konsequenzen für die Chemielehrer*innenbildung
ISBN 978-3-8325-5343-2 63.00 EUR
- 318 Laurence Müller (2021): Alltagsentscheidungen für den Chemieunterricht erkennen
und Entscheidungsprozesse explorativ begleiten
ISBN 978-3-8325-5379-1 59.00 EUR
- 319 Lars Ehlert (2021): Entwicklung und Evaluation einer Lehrkräftefortbildung zur Pla-
nung von selbstgesteuerten Experimenten
ISBN 978-3-8325-5393-7 41.50 EUR (open access)

- 320 Florian Seiler (2021): Entwicklung und Evaluation eines Seminarkonzepts zur Förderung der experimentellen Planungskompetenz von Lehramtsstudierenden im Fach Chemie
ISBN 978-3-8325-5397-5 47.50 EUR (open access)
- 321 Nadine Boele (2021): Entwicklung eines Messinstruments zur Erfassung der professionellen Unterrichtswahrnehmung von (angehenden) Chemielehrkräften hinsichtlich der Lernunterstützung
ISBN 978-3-8325-5402-6 46.50 EUR
- 322 Franziska Zimmermann (2022): Entwicklung und Evaluation digitalisierungsbezogener Kompetenzen von angehenden Chemielehrkräften
ISBN 978-3-8325-5410-1 49.50 EUR
- 323 Lars-Frederik Weiß (2021): Der Flipped Classroom in der Physik-Lehre. Empirische Untersuchungen in Schule und Hochschule
ISBN 978-3-8325-5418-7 51.00 EUR
- 324 Tilmann Steinmetz (2021): Kumulatives Lehren und Lernen im Lehramtsstudium Physik. Theorie und Evaluation eines Lehrkonzepts
ISBN 978-3-8325-5421-7 51.50 EUR
- 325 Kübra Nur Celik (2022): Entwicklung von chemischem Fachwissen in der Sekundarstufe I. Validierung einer Learning Progression für die Basiskonzepte „Struktur der Materie“, „Chemische Reaktion“ und „Energie“ im Kompetenzbereich „Umgang mit Fachwissen“
ISBN 978-3-8325-5431-6 55.00 EUR
- 326 Matthias Ungermann (2022): Förderung des Verständnisses von Nature of Science und der experimentellen Kompetenz im Schüler*innen-Labor Physik in Abgrenzung zum Regelunterricht
ISBN 978-3-8325-5442-2 55.50 EUR
- 327 Christoph Hoyer (2022): Multimedial unterstütztes Experimentieren im webbasierten Labor zur Messung, Visualisierung und Analyse des Feldes eines Permanentmagneten
ISBN 978-3-8325-5453-8 45.00 EUR
- 328 Tobias Schüttler (2022): Schülerlabore als interessefördernde authentische Lernorte für den naturwissenschaftlichen Unterricht nutzen
ISBN 978-3-8325-5454-5 50.50 EUR
- 329 Christopher Kurth (2022): Die Kompetenz von Studierenden, Schülerschwierigkeiten beim eigenständigen Experimentieren zu diagnostizieren
ISBN 978-3-8325-5457-6 58.50 EUR
- 330 Dagmar Michna (2022): Inklusiver Anfangsunterricht Chemie. Entwicklung und Evaluation einer Unterrichtseinheit zur Einführung der chemischen Reaktion
ISBN 978-3-8325-5463-7 49.50 EUR
- 331 Marco Seiter (2022): Die Bedeutung der Elementarisierung für den Erfolg von Mechanikunterricht in der Sekundarstufe I
ISBN 978-3-8325-5471-2 66.00 EUR

- 332 Jörn Hägele (2022): Kompetenzaufbau zum experimentbezogenen Denken und Arbeiten. Videobasierte Analysen zu Aktivitäten und Vorstellungen von Schülerinnen und Schülern der gymnasialen Oberstufe bei der Bearbeitung von fachmethodischer Instruktion
ISBN 978-3-8325-5476-7 56.50 EUR (open access)
- 333 Erik Heine (2022): Wissenschaftliche Kontroversen im Physikunterricht. Explorationsstudie zum Umgang von Physiklehrkräften und Physiklehramtsstudierenden mit einer wissenschaftlichen Kontroverse am Beispiel der Masse in der Speziellen Relativitätstheorie
ISBN 978-3-8325-5478-1 48.50 EUR (open access)
- 334 Simon Goertz (2022): Module und Lernzirkel der Plattform FLexKom zur Förderung experimenteller Kompetenzen in der Schulpraxis. Verlauf und Ergebnisse einer Design-Based Research Studie
ISBN 978-3-8325-5494-1 66.50 EUR
- 335 Christina Toschka (2022): Lernen mit Modellexperimenten. Empirische Untersuchung der Wahrnehmung und des Denkens in Analogien beim Umgang mit Modellexperimenten
ISBN 978-3-8325-5495-8 50.00 EUR (open access)
- 336 Alina Behrendt (2022): Chemiebezogene Kompetenzen in der Übergangsphase zwischen dem Sachunterricht der Primarstufe und dem Chemieunterricht der Sekundarstufe I
ISBN 978-3-8325-5498-9 40.50 EUR (open access)
- 337 Manuel Daiber (2022): Entwicklung eines Lehrkonzepts für eine elementare Quantenmechanik. Formuliert mit In-Out Symbolen
ISBN 978-3-8325-5507-8 48.50 EUR
- 338 Felix Pawlak (2022): Das Gemeinsame Experimentieren (an-)leiten. Eine qualitative Studie zum chemiespezifischen Classroom-Management
ISBN 978-3-8325-5508-5 46.50 EUR
- 339 Liza Dopatka (2022): Konzeption und Evaluation eines kontextstrukturierten Unterrichtskonzeptes für den Anfangs-Elektrizitätslehreunterricht
ISBN 978-3-8325-5514-6 69.50 EUR
- 340 Arne Bewersdorff (2022): Untersuchung der Effektivität zweier Fortbildungsformate zum Experimentieren mit dem Fokus auf das Unterrichtshandeln
ISBN 978-3-8325-5522-1 39.00 EUR (open access)
- 341 Thomas Christoph Münster (2022): Wie diagnostizieren Studierende des Lehramtes physikbezogene Lernprozesse von Schüler*innen?. Eine Videostudie zur Mechanik
ISBN 978-3-8325-5534-4 44.50 EUR (open access)
- 342 Ines Komor (2022): Förderung des symbolisch-mathematischen Modellverständnisses in der Physikalischen Chemie
ISBN 978-3-8325-5546-7 46.50 EUR

- 343 Verena Petermann (2022): Überzeugungen von Lehrkräften zum Lehren und Lernen von Fachinhalten und Fachmethoden und deren Beziehung zu unterrichtsnahem Handeln
ISBN 978-3-8325-5545-0 47.00 EUR (open access)
- 344 Jana Heinze (2022): Einfluss der sprachlichen Konzeption auf die Einschätzung der Qualität instruktionaler Unterrichtserklärungen im Fach Physik
ISBN 978-3-8325-5553-5 42.00 EUR (open access)
- 345 Jannis Weber (2022): Mathematische Modellbildung und Videoanalyse zum Lernen der Newtonschen Dynamik im Vergleich
ISBN 978-3-8325-5566-5 68.00 EUR (open access)
- 346 Fabian Sterzing (2022): Zur Lernwirksamkeit von Erklärvideos in der Physik. Eine Untersuchung in Abhängigkeit von ihrer fachdidaktischen Qualität und ihrem Einbettungsformat
ISBN 978-3-8325-5576-4 52.00 EUR (open access)
- 347 Lars Greitemann (2022): Wirkung des Tablet-Einsatzes im Chemieunterricht der Sekundarstufe I unter besonderer Berücksichtigung von Wissensvermittlung und Wissenssicherung
ISBN 978-3-8325-5580-1 50.00 EUR
- 348 Fabian Poensgen (2022): Diagnose experimenteller Kompetenzen in der laborpraktischen Chemielehrer*innenbildung
ISBN 978-3-8325-5587-0 48.00 EUR
- 349 William Lindlahr (2023): Virtual-Reality-Experimente. Entwicklung und Evaluation eines Konzepts für den forschend-entwickelnden Physikunterricht mit digitalen Medien
ISBN 978-3-8325-5595-5 49.00 EUR
- 350 Bert Schlüter (2023): Teilnahmemotivation und situationales Interesse von Kindern und Eltern im experimentellen Lernsetting KEMIE
ISBN 978-3-8325-5598-6 43.00 EUR
- 351 Katharina Nave (2023): Charakterisierung situativer mentaler Modellkomponenten in der Chemie und die Bildung von Hypothesen. Eine qualitative Studie zur Operationalisierung mentaler Modellkomponenten für den Fachbereich Chemie
ISBN 978-3-8325-5599-3 43.00 EUR
- 352 Anna B. Bauer (2023): Experimentelle Kompetenz Physikstudierender. Entwicklung und erste Erprobung eines performanzorientierten Kompetenzstrukturmodells unter Nutzung qualitativer Methoden
ISBN 978-3-8325-5625-9 47.00 EUR (open access)
- 353 Jan Schröder (2023): Entwicklung eines Performanztests zur Messung der Fähigkeit zur Unterrichtsplanung bei Lehramtsstudierenden im Fach Physik
ISBN 978-3-8325-5655-6 46.50 EUR
- 354 Susanne Gerlach (2023): Aspekte einer Fachdidaktik Körperpflege. Ein Beitrag zur Standardentwicklung
ISBN 978-3-8325-5659-4 45.00 EUR

- 355 Livia Murer (2023): Diagnose experimenteller Kompetenzen beim praktisch-naturwissenschaftlichen Arbeiten. Vergleich verschiedener Methoden und kognitive Validierung eines Testverfahrens
ISBN 978-3-8325-5657-0 41.50 EUR (open access)
- 356 Andrea Maria Schmid (2023): Authentische Kontexte für MINT-Lernumgebungen. Eine zweiteilige Interventionsstudie in den Fachdidaktiken Physik und Technik
ISBN 978-3-8325-5605-1 57.00 EUR (open access)
- 357 Julia Ortmann (2023): Bedeutung und Förderung von Kompetenzen zum naturwissenschaftlichen Denken und Arbeiten in universitären Praktika
ISBN 978-3-8325-5670-9 37.00 EUR (open access)
- 358 Axel-Thilo Prokop (2023): Entwicklung eines Lehr-Lern-Labors zum Thema Radioaktivität. Eine didaktische Rekonstruktion
ISBN 978-3-8325-5671-6 49.50 EUR
- 359 Timo Hackemann (2023): Textverständlichkeit sprachlich variierten physikbezogener Sachtexte
ISBN 978-3-8325-5675-4 41.50 EUR (open access)
- 360 Dennis Dietz (2023): Vernetztes Lernen im fächerdifferenzierten und integrierten naturwissenschaftlichen Unterricht aufgezeigt am Basiskonzept Energie. Eine Studie zur Analyse der Wirksamkeit der Konzeption und Implementation eines schulinternen Curriculums für das Unterrichtsfach „Integrierte Naturwissenschaften 7/8“
ISBN 978-3-8325-5676-1 49.50 EUR
- 361 Ann-Katrin Krebs (2023): Vielfalt im Physikunterricht. Zur Wirkung von Lehrkräftefortbildungen unter Diversitätsaspekten
ISBN 978-3-8325-5672-3 65.50 EUR (open access)
- 362 Simon Kaulhausen (2023): Strukturelle Ursachen für Klausurmisserfolg in Allgemeiner Chemie an der Universität
ISBN 978-3-8325-5699-0 37.50 EUR (open access)
- 363 Julia Eckoldt (2023): Den (Sach-)Unterricht öffnen. Selbstkompetenzen und motivationale Orientierungen von Lehrkräften bei der Implementation einer Innovation untersucht am Beispiel des Freien Explorierens und Experimentierens
ISBN 978-3-8325-5663-1 48.50 EUR (open access)
- 364 Albert Teichrow (2023): Physikalische Modellbildung mit dynamischen Modellen
ISBN 978-3-8325-5710-2 58.50 EUR (open access)
- 365 Sascha Neff (2023): Transfer digitaler Innovationen in die Schulpraxis. Eine explorative Untersuchung zur Förderung der Implementation
ISBN 978-3-8325-5687-7 59.00 EUR (open access)
- 366 Rahel Schmid (2023): Verständnis von Nature of Science-Aspekten und Umgang mit Fehlern von Schüler*innen der Sekundarstufe I. Am Beispiel von digital-basierten Lernprozessen im informellen Lernsetting Smartfeld
ISBN 978-3-8325-5722-5 53.50 EUR (open access)

- 367 Dennis Kirstein (2023): Individuelle Bedingungs- und Risikofaktoren für erfolgreiche Lernprozesse mit kooperativen Experimentieraufgaben im Chemieunterricht. Eine Untersuchung zum Zusammenhang von Lernvoraussetzungen, Lerntätigkeiten, Schwierigkeiten und Lernerfolg beim Experimentieren in Kleingruppen der Sekundarstufe I
ISBN 978-3-8325-5729-4 50.50 EUR (open access)
- 368 Frauke Düwel (2024): Argumentationslinien in Lehr-Lernkontexten. Potenziale englischer Fachtexte zur Chromatografie und deren hochschuldidaktische Einbindung
ISBN 978-3-8325-5731-7 63.00 EUR (open access)
- 369 Fabien Güth (2023): Interessenbasierte Differenzierung mithilfe systematisch variierten Kontextaufgaben im Fach Chemie
ISBN 978-3-8325-5737-9 48.00 EUR (open access)
- 370 Oliver Grewe (2023): Förderung der professionellen Unterrichtswahrnehmung und Selbstwirksamkeitsüberzeugungen hinsichtlich sprachsensibler Maßnahmen im naturwissenschaftlichen Sachunterricht. Konzeption und Evaluation einer video- und praxisbasierten Lehrveranstaltung im Masterstudium
ISBN 978-3-8325-5738-6 44.50 EUR (open access)
- 371 Anna Nowak (2023): Untersuchung der Qualität von Selbstreflexionstexten zum Physikunterricht. Entwicklung des Reflexionsmodells REIZ
ISBN 978-3-8325-5739-3 59.00 EUR (open access)
- 372 Dominique Angela Holland (2023): Bildung für nachhaltige Entwicklung (BNE) kooperativ gestalten. Vergleich monodisziplinärer und interdisziplinärer Kooperation von Lehramtsstudierenden bei der Planung, Durchführung und Reflexion von Online-BNE-Unterricht
ISBN 978-3-8325-5760-7 47.00 EUR (open access)
- 373 Renan Marcello Vairo Nunes (2024): MINT-Personal an Schulen. Eine Untersuchung der Arbeitssituation und professionellen Kompetenzen von MINT-Lehrkräften verschiedener Ausbildungswege
ISBN 978-3-8325-5778-2 51.00 EUR (open access)
- 374 Mats Kieserling (2024): Digitalisierung im Chemieunterricht. Entwicklung und Evaluation einer experimentellen digitalen Lernumgebung mit universeller Zugänglichkeit
ISBN 978-3-8325-5786-7 45.50 EUR
- 375 Cem Aydin Salim (2024): Die Untersuchung adaptiver Lernsettings im Themenbereich „Schwimmen und Sinken“ im naturwissenschaftlichen Unterricht
ISBN 978-3-8325-5787-4 49.00 EUR (open access)
- 376 Novid Ghassemi (2024): Evaluation eines Lehramtsmasterstudiengangs mit dem Profil Quereinstieg im Fach Physik. Erkenntnisse zu Eingangsbedingungen, professionellen Kompetenzen und Aspekten individueller Angebotsnutzung
ISBN 978-3-8325-5789-8 41.50 EUR (open access)

- 377 Martina Flurina Cavelti (2024): Entwicklung und Validierung eines Messinstruments zur Erfassung der Schülerkompetenzen im Bereich des wissenschaftlichen Skizzierens im Fach Chemie in der Sekundarstufe I
ISBN 978-3-8325-5829-1 44.00 EUR (open access)
- 378 Tom Bleckmann (2024): Formatives Assessment auf Basis von maschinellem Lernen. Eine Studie über automatisiertes Feedback zu Concept Maps aus dem Bereich Mechanik
ISBN 978-3-8325-5842-0 45.00 EUR (open access)

Vollständige Übersicht unter: <https://www.logos-verlag.de/spcl>

Alle erschienenen Bücher können unter der angegebenen ISBN direkt online (<http://www.logos-verlag.de>) oder telefonisch (030 - 42 85 10 90) beim Logos Verlag Berlin bestellt werden.

Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf und Mathias Ropohl

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

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

Kontaktadressen:

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

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

Wissenschaftliches Skizzieren bezeichnet die Fähigkeiten und Fertigkeiten während des Experimentierprozesses objektive Skizzen anzufertigen, die als Basis für Erkenntnisgewinnung dienen. Diese Handlungskompetenz ist in der Schulpraxis relevant für eine nicht-verbale Darstellung von Experimentierprozessen. Es fehlt bisher jedoch ein validiertes Kompetenzmodell.

Ziel dieser Arbeit ist es, ein Kompetenzmodell und ein Messinstrument zur Erfassung des wissenschaftlichen Skizzierens zu entwickeln und zu validieren. Es wird eine Skala zum wissenschaftlichen Skizzieren mit 9 Items entwickelt und mit $N = 374$ Schüler:innen (7./9. Jahrgangstufe) und den Variablen Kognition, Strategiewissen beim Experimentieren, Lesefähigkeit und spezifisches Fach- und Methodenwissen validiert.

Die Ergebnisse zeigen zunächst, dass sechs Items für ein bezüglich Testgüte valides sowie reliables Messinstrument genügen. Die externe Validierung zeigt, dass der Test zur Kompetenz des wissenschaftlichen Skizzierens erwartungsgemäß ein anderes Konstrukt erhebt als die Kognition, das Strategiewissen zum Experimentieren oder die Lesefähigkeit. Das entwickelte Messinstrument zur Erfassung des wissenschaftlichen Skizzierens ist somit für diagnostische Zwecke einsetzbar. Es kann als ein Messinstrument zur summativen Beurteilung in der Schulpraxis weiterentwickelt werden.

Logos Verlag Berlin

ISBN 978-3-8325-5829-1